UNIVERSITY OF DERBY

Smart City: A Traffic Signal Control System for Reducing the Effects of Traffic Congestion in Urban Environments

James Hardy

Doctor of Philosophy 2019

Table of Contents

[List of figures vii](#_Toc9527683)

[List of tables xi](#_Toc9527684)

[Acknowledgments xiii](#_Toc9527685)

[Abstract xiv](#_Toc9527686)

[Abbreviations xvi](#_Toc9527687)

[Publications xviii](#_Toc9527688)

[Chapter 1: Introduction 1](#_Toc9527689)

[1.1 Background 1](#_Toc9527690)

[1.2 Problem Statement 2](#_Toc9527691)

[1.3 Aim and Objectives 4](#_Toc9527692)

[1.3.1 Aim 4](#_Toc9527693)

[1.3.2 Objectives 4](#_Toc9527694)

[1.4 Research Justification 6](#_Toc9527695)

[1.5 Research Novelty 10](#_Toc9527696)

[1.6 Methodology 13](#_Toc9527697)

[1.6.1 Research Methodology 13](#_Toc9527698)

[1.6.2 Project Methodology 15](#_Toc9527699)

[1.7 Purpose of Research 15](#_Toc9527700)

[1.8 Scope of Work 20](#_Toc9527701)

[1.9 Thesis Structure 24](#_Toc9527702)

[Chapter 2: Literature Review 26](#_Toc9527703)

[2.1 Introduction 26](#_Toc9527704)

[2.2 Related Works 27](#_Toc9527705)

[2.3 Traffic Light Control Systems 30](#_Toc9527706)

[2.3.1 Outline 30](#_Toc9527707)

[2.3.2 Sequencing 30](#_Toc9527708)

[2.3.3 Detectors 32](#_Toc9527709)

[2.3.4 Control 33](#_Toc9527710)

[2.3.5 Predictive Methods 45](#_Toc9527711)

[2.4.1 Chaos Theory 49](#_Toc9527712)

[2.4.2 Queueing Theory 51](#_Toc9527713)

[2.4.3 Genetic Algorithms 54](#_Toc9527714)

[2.4.4 Cellular Automata 56](#_Toc9527715)

[2.4.5 Petrinets 56](#_Toc9527716)

[2.4.6 Game Theory 58](#_Toc9527717)

[2.4.7 Simulation 61](#_Toc9527718)

[2.4.8 Human Interaction and Stochastic Flow 63](#_Toc9527719)

[2.4.9 Limitations Of Individual Optimisations 64](#_Toc9527720)

[Chapter 3: Congestion Modelling 67](#_Toc9527721)

[3.1 Outline 67](#_Toc9527722)

[3.2 Congestion Effects 68](#_Toc9527723)

[3.2.1 Negative Aspects of Congestion 68](#_Toc9527724)

[3.2.2 The Economic Benefit of Congestion 70](#_Toc9527725)

[3.2.3 Challenges of Congestion 72](#_Toc9527726)

[3.3 Quantifying Congestion 73](#_Toc9527727)

[3.3.1 Road Capacity 74](#_Toc9527728)

[3.3.2 PCU (PCE) values 79](#_Toc9527729)

[3.3.3 Traffic Density 80](#_Toc9527730)

[3.3.4 Capacity Drop 81](#_Toc9527731)

[3.3.5 Braking Distance 82](#_Toc9527732)

[3.3.6 Fundamental Diagrams 85](#_Toc9527733)

[3.4 How Isolated Traffic Lights Cause Delays 86](#_Toc9527734)

[3.4.1 Delay Periods at Traffic Lights 93](#_Toc9527735)

[3.4.2 Micro, Macro and Mesoscopic Flow 95](#_Toc9527736)

[3.4.3 Causes of Congestion 99](#_Toc9527737)

[3.4.4 Congestion Taxonomy 100](#_Toc9527738)

[3.4.5 General Congestion Resolution 101](#_Toc9527739)

[Chapter 4: Traffic Modelling 109](#_Toc9527740)

[4.1 Traffic Presence 109](#_Toc9527741)

[4.1.1 Traffic Efficiency 109](#_Toc9527742)

[4.2 Traffic Flow 112](#_Toc9527743)

[4.3 Traffic Types 117](#_Toc9527744)

[4.4 Converging and Diverging Traffic 118](#_Toc9527745)

[4.5 Why the Bus (or Train) is Always Late 126](#_Toc9527746)

[Chapter 5: Detail of Conceptual Solution 131](#_Toc9527747)

[5.1 Outline 131](#_Toc9527748)

[5.2 Controller Concept 132](#_Toc9527749)

[5.3 AFRC Proposal 134](#_Toc9527750)

[5.3.1 General Description 134](#_Toc9527751)

[5.3.2 Benefit of AFRC 136](#_Toc9527752)

[5.3.3 Changes Required To Accommodate Practical AFRC 138](#_Toc9527753)

[5.3.4 Future Need for Traffic Control 139](#_Toc9527754)

[5.4 Algorithm Design 140](#_Toc9527755)

[5.4.1 Introduction 140](#_Toc9527756)

[5.4.2 Testing 143](#_Toc9527757)

[Chapter 6: Simulations and Results 162](#_Toc9527758)

[6.1 Section Outline 162](#_Toc9527759)

[6.2 Validity of Simulations 162](#_Toc9527760)

[6.3 Choice of Simulator 166](#_Toc9527761)

[6.4 Simulation Host System 167](#_Toc9527762)

[6.4.1 Hardware: 167](#_Toc9527763)

[6.4.2 Software 167](#_Toc9527764)

[6.5 General Setup of Experiments 168](#_Toc9527765)

[6.5.1 Experiment Description Tables 169](#_Toc9527766)

[6.6 Experiment 1: 5x4 Grid 170](#_Toc9527767)

[6.6.1 Results: 172](#_Toc9527768)

[6.6.2 Summary 174](#_Toc9527769)

[6.7 Experiment 2: 7x6 Grid 175](#_Toc9527770)

[6.7.1 Setup 175](#_Toc9527771)

[6.7.2 Results 178](#_Toc9527772)

[6.7.3 Summary 180](#_Toc9527773)

[6.8 Experiment 3: 7x6 realflow grid 180](#_Toc9527774)

[6.8.1 Setup 180](#_Toc9527775)

[6.8.2 Results 185](#_Toc9527776)

[6.8.3 Summary 189](#_Toc9527777)

[6.9 Experiments 4-9: 7x6 realflow grid 190](#_Toc9527778)

[6.9.1 Setup 190](#_Toc9527779)

[6.9.2 Experiment 4 Summary: 193](#_Toc9527780)

[6.9.3 Experiment 4 Results 195](#_Toc9527781)

[6.9.4 Experiment 4 interpretation and conclusions 196](#_Toc9527782)

[6.9.5 Experiment 5 Results 200](#_Toc9527783)

[6.9.6 Experiment 6 Results 201](#_Toc9527784)

[6.9.7 Experiment 7 Results 202](#_Toc9527785)

[6.9.8 Experiment 8 Results 203](#_Toc9527786)

[6.9.9 Experiment 9 Results 204](#_Toc9527787)

[6.9.10 Experiment 5 – 9 Summary 205](#_Toc9527788)

[6.10 Experiment 10: Multidirectional Flow 205](#_Toc9527789)

[6.10.1 Setup 206](#_Toc9527790)

[6.10.2 Results 209](#_Toc9527791)

[6.10.3 Summary 214](#_Toc9527792)

[Chapter 7: Conclusions and Future works 215](#_Toc9527793)

[7.1 Conclusions 215](#_Toc9527794)

[7.2 Limitations of AFRC 216](#_Toc9527795)

[7.3 Future Works 217](#_Toc9527796)

[7.3.1 Autonomous Vehicles 217](#_Toc9527797)

[7.3.2 Connected Vehicle Priority 219](#_Toc9527798)

[7.3.3 Reduced Headway and Reaction Speed 220](#_Toc9527799)

[7.3.4 Connected Vehicle Platooning 221](#_Toc9527800)

[7.3.5 Road Pricing, Cost of Ownership 222](#_Toc9527801)

[7.3.6 Shared Ownership, Pay-Per-Use 223](#_Toc9527802)

[7.3.7 Big Data and Traffic Predictions 225](#_Toc9527803)

[7.4 Potential Research Fields 226](#_Toc9527804)

[7.4.1 Connected and Autonomous Vehicles 227](#_Toc9527805)

[7.4.2 Traffic Signals Communication 227](#_Toc9527806)

[7.4.3 Prediction Methods 227](#_Toc9527807)

[7.4.4 Economics and the need for change 228](#_Toc9527808)

[References 230](#_Toc9527809)

[Appendix 1: Traffic Light Control Algorithms 245](#_Toc9527810)

[Timed Algorithm 245](#_Toc9527811)

[Modified VA Algorithm 247](#_Toc9527812)

[Traditional VA Algorithm 249](#_Toc9527813)

[AFRC Algorithm 251](#_Toc9527814)

[Appendix 2: Traffic signal terminology 254](#_Toc9527815)

[Appendix 3: Comparative safety of road classes 257](#_Toc9527816)

[Appendix 4: Visit to Parliament 260](#_Toc9527817)

[Appendix 5: Autonomous vehicle testing visit 261](#_Toc9527818)

[Appendix 6: Best Paper Certificate 262](#_Toc9527819)

[Appendix 7: Sample database outputs 263](#_Toc9527820)

List of figures

[Figure 1.1: Fast Moving Traffic Preventing Access 8](#_Toc9527821)

[Figure 1.2: Potential for Gridlocking 8](#_Toc9527822)

[Figure 1.3: “Derby Junctions” Road Map from Openstreetmap.org 17](#_Toc9527823)

[Figure 1.4: Junction Directions 23](#_Toc9527824)

[Figure 2.1: Simple Crossroad Junction with Signal Directions 31](#_Toc9527825)

[Figure 2.2: MOVA Detector Placing. From MCH1642 iss. C 35](#_Toc9527826)

[Figure 2.3: Signal Coordination across Six Junctions 38](#_Toc9527827)

[Figure 2.4: Components of Queueing Theory 51](#_Toc9527828)

[Figure 2.5: Queueing at a Non-isolated Junction 52](#_Toc9527829)

[Figure 2.6: Queueing Theory Applied to Non-isolated Crossroad Junction 54](#_Toc9527830)

[Figure 2.7: Example Timed Petrinet of Single Traffic Light Head 57](#_Toc9527831)

[Figure 3.1: Occupancy vs Headway 76](#_Toc9527832)

[Figure 3.2: Theoretical Lane Capacity with 2 Second Gap 77](#_Toc9527833)

[Figure 3.3: Vehicle Density per Mile At Various Speeds 81](#_Toc9527834)

[Figure 3.4: Fundamental Diagrams after Greenshields [90] 85](#_Toc9527835)

[Figure 3.5: Fundamental Diagrams after Smulders [91] 85](#_Toc9527836)

[Figure 3.6: Density/Flow from del Castillo Relating Flow State to Critical Value [92] 86](#_Toc9527837)

[Figure 3.7: Single, One-Way, 30 Mph Lane at Full Capacity of 1800pcu/hr 87](#_Toc9527838)

[Figure 3.8: Single Lane with Controlled Crossroad 87](#_Toc9527839)

[Figure 3.9: Theoretical 50/50 split 88](#_Toc9527840)

[Figure 3.10: Practical Light Sequence 89](#_Toc9527841)

[Figure 3.11: Effect of Increasing Flow Rate; Pcu Flow as a Function of Cycle Time 93](#_Toc9527842)

[Figure 3.12: Queue length vs Cycles, Ignoring Transition Loss 97](#_Toc9527843)

[Figure 3.13: Queue length vs cycles, Including Transition Loss 98](#_Toc9527844)

[Figure 4.1: Concept of Tidal Flow 113](#_Toc9527845)

[Figure 4.2: Simple Crossroad Junction 114](#_Toc9527846)

[Figure 4.3: Traffic Converging at a Destination 120](#_Toc9527847)

[Figure 4.4: Traffic Converging from an Origin 120](#_Toc9527848)

[Figure 4.5: Overlay Concept 122](#_Toc9527849)

[Figure 4.6: Time Stacked Overlays 123](#_Toc9527850)

[Figure 4.7: "School run" Converging And Diverging Flow 125](#_Toc9527851)

[Figure 4.8:City overlay 1 - single O, D 126](#_Toc9527852)

[Figure 4.9: City overlay 2 - Multi O, Single D 126](#_Toc9527853)

[Figure 4.10: City Overlay 3 - Multi O, Single large D 126](#_Toc9527854)

[Figure 4.11: City Overlay - combined 126](#_Toc9527855)

[Figure 4.12: Velocity/Time Curve, Uniform Acceleration And Deceleration 128](#_Toc9527856)

[Figure 5.1: Approach Detection Control 132](#_Toc9527857)

[Figure 5.2: Serial queueing with an output buffer space 133](#_Toc9527858)

[Figure 5.3: AFRC queueing control – stage 1 135](#_Toc9527859)

[Figure 5.4: AFRC with forward notification 136](#_Toc9527860)

[Figure 5.5: Simple VA Control 142](#_Toc9527861)

[Figure 5.6: VA tls with Exit Detection 142](#_Toc9527862)

[Figure 5.7: Aimsun Simulated Roadway 143](#_Toc9527863)

[Figure 5.8: Diagrammatically Simplified “Derby Junctions” 147](#_Toc9527864)

[Figure 5.9: Simple Three Junction Layout 149](#_Toc9527865)

[Figure 5.10: Junction Layout With Detectors 151](#_Toc9527866)

[Figure 5.11: Typical Upstream/Downstream Link 152](#_Toc9527867)

[Figure 5.12: 4x5 Traffic Grid 152](#_Toc9527868)

[Figure 5.13: 7x6 Grid, All Junctions Have Four Approaches 154](#_Toc9527869)

[Figure 5.14: 7x6 “Realflow” Grid 155](#_Toc9527870)

[Figure 5.15: Traffic Queueing to Enter From Commercial Area 156](#_Toc9527871)

[Figure 5.16: Modified Commercial Area Link with Additional Detectors 156](#_Toc9527872)

[Figure 6.1: example description table 170](#_Toc9527873)

[Figure 6.2: Experiment 1 Time Delay across Network 172](#_Toc9527874)

[Figure 6.3: 7x6 Grid Repeated for Clarity 176](#_Toc9527875)

[Figure 6.4: Experiment 2 Time Delay across Network 178](#_Toc9527876)

[Figure 6.5: 7x6 Realfow Grid 183](#_Toc9527877)

[Figure 6.6: Experiment 3 Time Delay across Network 185](#_Toc9527878)

[Figure 6.7: Experiment 3 Traffic Input Flow 186](#_Toc9527879)

[Figure 6.8: Experiment 3 Vehicles Waiting to Enter 188](#_Toc9527880)

[Figure 6.9: 7x6 Realflow Grid - Experiments 4 - 9 191](#_Toc9527881)

[Figure 6.10: Experiment 4 Delay across Network 195](#_Toc9527882)

[Figure 6.11: Experiment 4 Stop Time 195](#_Toc9527883)

[Figure 6.12: Experiment 4 Travel Time 195](#_Toc9527884)

[Figure 6.13: Experiment 4 Vehicles Inside Network 195](#_Toc9527885)

[Figure 6.14: Experiment 4 Vehicles Waiting to Enter 195](#_Toc9527886)

[Figure 6.15: Experiment 4 Travel Distance 195](#_Toc9527887)

[Figure 6.16: Experiment 4 Showing Traffic Input Volume 197](#_Toc9527888)

[Figure 6.17: Experiment 4 AFRC Network Efficiency 198](#_Toc9527889)

[Figure 6.18: Experiment 4 VA Network Efficiency 199](#_Toc9527890)

[Figure 6.19: Experiment 5 Delay across Network 200](#_Toc9527891)

[Figure 6.20: Experiment 5 Stop Time 200](#_Toc9527892)

[Figure 6.21: Experiment 5 Travel Time 200](#_Toc9527893)

[Figure 6.22: Experiment 5 Vehicles Inside Network 200](#_Toc9527894)

[Figure 6.23: Experiment 5 Vehicles Waiting to Enter 200](#_Toc9527895)

[Figure 6.24: Experiment 5 Travel Distance 200](#_Toc9527896)

[Figure 6.25: Experiment 6 Delay across Network 201](#_Toc9527897)

[Figure 6.26: Experiment 6 Stop Time 201](#_Toc9527898)

[Figure 6.27: Experiment 6 Travel Time 201](#_Toc9527899)

[Figure 6.28: Experiment 6 Vehicles Inside Network 201](#_Toc9527900)

[Figure 6.29: Experiment 6 Vehicles Waiting to Enter 201](#_Toc9527901)

[Figure 6.30: Experiment 6 Travel Distance 201](#_Toc9527902)

[Figure 6.31: Experiment 7 Delay across Network 202](#_Toc9527903)

[Figure 6.32: Experiment 7 Stop Time 202](#_Toc9527904)

[Figure 6.33: Experiment 7 Travel Time 202](#_Toc9527905)

[Figure 6.34: Experiment 7 Vehicles Inside Network 202](#_Toc9527906)

[Figure 6.35: Experiment 7 Vehicles Waiting to Enter 202](#_Toc9527907)

[Figure 6.36: Experiment 7 Travel Distance 202](#_Toc9527908)

[Figure 6.37: Experiment 8 Delay across Network 203](#_Toc9527909)

[Figure 6.38: Experiment 8 Stop Time 203](#_Toc9527910)

[Figure 6.39: Experiment 8 Travel Time 203](#_Toc9527911)

[Figure 6.40: Experiment 8 Vehicles Inside Network 203](#_Toc9527912)

[Figure 6.41: Experiment 8 Vehicles Waiting to Enter 203](#_Toc9527913)

[Figure 6.42: Experiment 8 Travel Distance 203](#_Toc9527914)

[Figure 6.43: Experiment 9 Delay across Network 204](#_Toc9527915)

[Figure 6.44: Experiment 9 Stop Time 204](#_Toc9527916)

[Figure 6.45: Experiment 9 Travel Time 204](#_Toc9527917)

[Figure 6.46: Experiment 9 Vehicles Inside Network 204](#_Toc9527918)

[Figure 6.47: Experiment 9 Vehicles Waiting to Enter 204](#_Toc9527919)

[Figure 6.48: Experiment 9 Travel Distance 204](#_Toc9527920)

[Figure 6.49: 7x6 Grid with Detector Update 206](#_Toc9527921)

[Figure 6.50: Experiment 10 Time Delay across Network 211](#_Toc9527922)

[Figure 6.51: Experiment 10 Absolute Traffic Profile 211](#_Toc9527923)

[Figure 6.52: Experiment 3 - 9 Traffic Profile 212](#_Toc9527924)

[Figure 6.53: Experiment 10 Commercial Area Traffic Flow 212](#_Toc9527925)

[Figure 6.54: Experiment 10 Vehicles Waiting to Enter Network 213](#_Toc9527926)

[Figure 6.55: Experiment 10 Vehicles Inside Network 213](#_Toc9527927)

[Figure 6.56: Experiment 10 Total Travel Distance 214](#_Toc9527928)

[Figure 7.1: Google Image of Derby Railway Station 225](#_Toc9527929)

[Figure 7.2: Google Image of “The Spot” Derby 225](#_Toc9527930)

List of tables

[Table 1.1: Google Prediction Compared to Actual Journey Time and Speed 19](#_Toc9527931)

[Table 2.1: Essential / Minimal Light Sequence 31](#_Toc9527932)

[Table 2.2: Light Sequence with Safety “All Red” Stages 31](#_Toc9527933)

[Table 2.3: Full Sequence for Simple Junction 32](#_Toc9527934)

[Table 2.4: Vehicle Flow Through Six Coordinated Signals 38](#_Toc9527935)

[Table 2.5: Reverse Flow at Coordinated Signals 39](#_Toc9527936)

[Table 3.1: Speed for Vehicles to Move Own Distance in 2 Seconds 76](#_Toc9527937)

[Table 3.2: Comparison of Capacity Based on Headway and Occupancy 77](#_Toc9527938)

[Table 3.3: Measured Capacity on Multilane Roads 78](#_Toc9527939)

[Table 3.4: Calculated PCU for Bus, Using Occupancy Value 80](#_Toc9527940)

[Table 3.5: Braking Distances taken from The Highway Code 83](#_Toc9527941)

[Table 3.6: Stopping Time Based on Highway Code Values 84](#_Toc9527942)

[Table 3.7: Stopping Times Based on 1G Deceleration 84](#_Toc9527943)

[Table 3.8: Split and Availability - 0 sec all red time 91](#_Toc9527944)

[Table 3.9: Split and Availability - 2 sec all red time 91](#_Toc9527945)

[Table 3.10: Opposing Split and Junction Availability 92](#_Toc9527946)

[Table 3.11: flow rates based on cycle times 93](#_Toc9527947)

[Table 3.12: Queue Statistics for Zero Transition Losses 97](#_Toc9527948)

[Table 3.13: Queue Statistics with 2.5 PCU Transition Losses 98](#_Toc9527949)

[Table 3.14: Extended Stages of Action Determination 107](#_Toc9527950)

[Table 4.1: Possible Junction Sequences 114](#_Toc9527951)

[Table 4.2: Unequal Time Allocation to Improve Tidal Flow 115](#_Toc9527952)

[Table 4.3: Variety of Common Transport Methods Observable on Urban Roads 118](#_Toc9527953)

[Table 4.4: Average And Max Velocity Per Stage 130](#_Toc9527954)

[Table 4.5: Ave And Max Velocity To Recover Previous Stage Delay 130](#_Toc9527955)

[Table 5.1: Example Traffic Flow Profile 157](#_Toc9527956)

[Table 5.2: Example Traffic Demand 157](#_Toc9527957)

[Table 6.1: Experiment 1 Traffic Flow - Normal 170](#_Toc9527958)

[Table 6.2: Experiment 1 Traffic Flow - Busy 171](#_Toc9527959)

[Table 6.3: Experiment 1 Traffic Flow - Quiet 171](#_Toc9527960)

[Table 6.4: Experiment 1 Information Summary 171](#_Toc9527961)

[Table 6.5: Experiment 1 Results - Time Delay (sec) 173](#_Toc9527962)

[Table 6.6: Experiment 2 Information Summary 176](#_Toc9527963)

[Table 6.7: Experiment 2 Traffic Flow - Normal 177](#_Toc9527964)

[Table 6.8: Experiment 2 Traffic Flow – Rushhour 177](#_Toc9527965)

[Table 6.9: Experiment 2 Traffic Flow - Quiet 177](#_Toc9527966)

[Table 6.10: Experiment 2 Traffic Flow - None 178](#_Toc9527967)

[Table 6.11: Experiment 2 Results - Time Delay (sec) 179](#_Toc9527968)

[Table 6.12: Experiment 3 Information Summary 183](#_Toc9527969)

[Table 6.13: Experiment 3 Traffic Flow - Normal 183](#_Toc9527970)

[Table 6.14: Experiment 3 Traffic Flow – Rushhour 184](#_Toc9527971)

[Table 6.15: Experiment 3 Traffic Flow - Quiet 184](#_Toc9527972)

[Table 6.16: Experiment 3 Traffic Flow - None 184](#_Toc9527973)

[Table 6.17: Experiment 3 Time Delay Across Network (sec) 186](#_Toc9527974)

[Table 6.18: Experiment 3 Traffic Input Flow (vehicles) 187](#_Toc9527975)

[Table 6.19: Experiment 3 Vehicles Waiting to Enter 188](#_Toc9527976)

[Table 6.20: Experiment 4-9 Information Summary 192](#_Toc9527977)

[Table 6.21: Experiment 4-9 Traffic Flow - Normal 192](#_Toc9527978)

[Table 6.22: Experiment 4-9 Traffic Flow – Rushhour 192](#_Toc9527979)

[Table 6.23: Experiment 4-9 Traffic Flow - Quiet 193](#_Toc9527980)

[Table 6.24: Experiment 4-9 Traffic Flow - None 193](#_Toc9527981)

[Table 6.25: Experiment 10 Information Summary 207](#_Toc9527982)

[Table 6.26: Experiment 10 Traffic Flow - Peak to Commercial 207](#_Toc9527983)

[Table 6.27: Experiment 10 Traffic Flow – Modest to-from Commercial 208](#_Toc9527984)

[Table 6.28: Experiment 10 Traffic Flow - Peak from Commercial 208](#_Toc9527985)

[Table 6.29: Experiment 10 Traffic Flow – Minimal Residential 208](#_Toc9527986)

[Table 6.30: Experiment 10 Traffic Flow - None 209](#_Toc9527987)

[Table 7.1: Stopping Times Based on 1G Deceleration 221](#_Toc9527988)

[Table A3.1: Accident Statistics for road types 2014 257](#_Toc9527989)

[Table A3.2: Incidents per mile 2014 258](#_Toc9527990)

[Table A3.3: Incidents per mile based on Traffic Estimates report 259](#_Toc9527991)

Acknowledgments

I would like to thank some of the most dedicated, inspirational and supportive people that I have been honoured to meet through the course of this work. I am truly humbled by their assistance.

To my wife and family for their understanding and tolerance. Thank you for listening and reading this work.

To Dr Stuart Berry for ideas, insights and supportive challenges of this work and concepts and also for the hours of proofreading. Dr Asma Adnane for discussion and motivation. Dr Virginia Franqueira for convincing me that I can get to the end.

To the people that I have met at conferences, seminars and trade events. The many people from ITS-UK, IMPart and the Transport Catapult and for the assistance and encouragement of Aimsun.

Finally to Prof Lu Liu for giving me this opportunity

Thank You

Abstract

This thesis addresses the detrimental effects of road traffic congestion in the Smart City environment. Urban congestion is a recognisable problem that affects much of the world’s population through delays and pollution although the delays are not an entirely modern phenomena. The progressive increase in urbanisation and the numbers of powered road vehicles have led to an increasing need to control traffic in order to maintain flows and avoid gridlock situations. Signalised methods typically control flows through reduction, frequently increasing delays, holding traffic within the urban area and increasing local pollution. The current levels of vehicular congestion may relate to an increase in traffic volumes of 300% over 50 years while traffic control methods based on delaying moving traffic have changed very little. Mobility and Socio-economics indicate that the number of active road vehicles will increase or at least remain at the same levels in the foreseeable future and as a result congestion will continue to be a problem. The Smart City concept is intended to improve the urban environment through the application of advanced technology. Within the context of road transportation, the urban area consists of a wide variety of low to moderate speed transportation systems ranging from pedestrians to heavy goods vehicles. Urban roadways have a large number of junctions where the transport systems and flows interact presenting additional and more complex challenges as compared to high speed dual carriageways and motorways. Congestion is a function of population density while car ownership is an indicator of affluence; road congestion can therefore be seen as an indicator of local economic and social prosperity. Congestion cannot be resolved while there is a social benefit to urbanisation, high density living and a materialistic population. Recognising that congestion cannot be resolved, this research proposes a method to reduce the undesirable consequences and side effects of traffic congestion such as transit delays, inefficient fuel use and chemical pollution without adversely affecting the social and economic benefits.

Existing traffic signal systems manage traffic flows based on traffic arrivals, prediction and traffic census models. Flow modification is accomplished by introducing delays through signal transition in order to prioritise a conflicting direction. It is incorrectly assumed that traffic will always be able to move and therefore signal changes will always have an effect. Signal transitions result in lost time at the junction. Existing Urban Traffic Control systems have limited capability as they are unable to adapt immediately to unexpected conditions, have a finite response, cannot modify stationary flow and may introduce needless losses through inefficient transition. This research proposes and develops Available Forward Road Capacity (AFRC), an algorithm with the ability to detect the onset of congestion, actively promote clearance, prevent unnecessary losses due to ineffective transitions and can influence other AFRC equipped junctions to ensure the most efficient use of unoccupied road capacity. AFRC is an additional function that can be applied to existing traffic controllers, becoming active only during congestion conditions; as a result it cannot increase congestion above current levels. By reducing the duration of congestion periods, AFRC reduces delays, improves the efficiency of fuel use and reduces pollution. AFRC is a scalable, multi-junction generalised solution which is able to manage traffic from multiple directions without prior tuning; it can detect and actively resolve problems with stationary traffic. AFRC is evaluated using a commercial traffic simulation system and is shown to resolve inbound and outbound congestion in less time than Vehicle Actuated and Fully Timed systems when simulating both morning and evening rush-hours.

Abbreviations

|  |  |
| --- | --- |
| ADSL | Asymmetric Digital Subscriber Lines |
| AFRC | Available Forward Road Capacity |
| AV | Autonomous Vehicles |
| bps | Bits per second |
| C-AV | Connected Autonomous Vehicles |
| CFP | cyclic flow profile |
| CV | Connected Vehicles |
| DegS | Degree of Saturation (commonly DS) |
| EIGRP | Enhanced Interior Gateway Routing Protocol |
| HGV | Heavy Goods Vehicle |
| HMI | Human Machine Interface |
| IGP | Interior Gateway Protocol |
| ITU | International Telecommunications Union |
| KSI | killed or seriously injured |
| LGV | Light Goods Vehicle |
| Lidar | Light Detection and Ranging / Light Imaging, Detection and Ranging  / portmanteau of Light and Radar |
| lpu | link profile units |
| MaaS | Mobility as a Service |
| OSPF | Open Shortest Path First |
| PID | Proportional, Integral, Derivative |
| POV | Privately Owned Vehicles |
| PSV | Passenger Service Vehicle (typically a bus) |
| RIP | Routing Information Protocol |
| SCATS | Sydney Coordinated Adaptive Traffic System |
| SCOOT | Split, Cycle and Offset Optimisation Technique |
| SMMT | Society of Motor Manufacturers and Traders |
| SOTL | Self Optimised Traffic Lights |
| TDM | Time Division Multiplexing |
| TTL | Time to Live |
| URL | Uniform Resource Locator |
| UTC | Urban Traffic Control |
| VLSM | Variable Length Subnet Masks |
| V2V | Vehicle to vehicle (communication) |
| V2i | Vehicle to infrastructure (communication) |
| V2x | Vehicle to anything (communication) |

Publications

* James Hardy, Lu Liu, Lei Cui, Jianxin Li; Internet-based Virtual Computing Infrastructure for Cloud Computing, Book Chapter for Principles, Methodologies, and Service-Oriented Approaches for Cloud Computing, 2013, pp371 - 389, IGI Global. ISBN 978-1-4666-2854-0
* James Hardy, Lu Liu, Nick Bessis, Lei Cui, Jianxin Li; Performance Measurements in iVIC Virtual Machine Clones, Proceedings of 26th International Conference on Advanced Information Networking and Applications Workshops, Fukuoka, Japan, March 26-29, 2012, pp. 1258 - 1263. IEEE Computer Society Press. DOI:10.1109/WAINA.2012.41
* James Hardy, Lu Liu, Nick Antonopoulos, Weining Liu, Lei Cui, Jianxin Li; Assessment and Evaluation of Internet-based Virtual Computing Infrastructure, Proceedings of 15th IEEE International Symposium on Object/component/service-oriented Real-time distributed Computing (ISORC 2012), Shenzhen, China, April 11-13, 2012. pp. 39-46, ISSN 1555-0885/12. IEEE Computer Society Press. DOI: 10.1109/ISORC.2012.14.
* Zhenxing Liu, Lu Liu, James Hardy, Ashiq Anjum, Richard Hill, Nick Antonopoulos; Dot-base62x: A Compact Textual Representation of IPv6 Address for Clouds, Proceedings of 4th IEEE International Conference on Utility and Cloud Computing (UCC 2011), December, Melbourne, Australia 2011. pp. 427 - 432, ISBN: 978-1-4577-2116-8. IEEE Computer Society Press. DOI: 10.1109/UCC.2011.70
* Zhenxing Liu, Lu Liu, James Hardy, Ashiq Anjum, Richard Hill, Nick Antonopoulos; Dot-base62x: Building A Compact and User-friendly Text Representation Scheme of IPv6 Addresses for Cloud Computing, Journal of Cloud Computing, Springer, Vol 1(3), 2012, ISSN: 2192-113X. DOI:10.1186/2192-113X-1-3
* Lei Cui, Bo Li, Jianxin Li, James Hardy, Lu Liu; Software Aging in Virtualized Environments: Detection and Prediction. IEEE 18th International Conference on Parallel and Distributed Systems (ICPADS), 2012: 718-719. DOI: 10.1109/ICPADS.2012.111
* M. Ahmed, Lu Liu, J. Hardy, Bo Yuan; An Efficient Algorithm for Partially Matched Web Services Based on Consumer’s QoS Requirements; 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing (UCC); Dec 8-11 2014; DOI: 10.1109/UCC.2014.140
* Zhenxing Liu, Hao Liu, James Hardy, Lu Liu; Uniform Resource Identifier Encoded by Base62x; IOP 2015, Beijing, August 10-14 2015
* James Hardy, Lu Liu; Future transport and the Internet of People; IOP 2015, Beijing, August 10-14 2015
* James Hardy; Chapter 13: Applying Queueing Theory to the design of a Traffic Light Controller. DOI: 10.1007/978-3-319-55417-4\_13. In book: Guide to Computational Modelling for Decision Processes - Theory, Algorithms, Techniques and Applications. Stuart Berry, Val Lowndes and Marcello Trovati. ISBN: 978-3-319-55416-7. Jan 2017
* John Panneerselvam, James Hardy, Lu Liu, Bo Yuan, Nick Antonopoulos; Mobilouds: An Energy Efficient MCC Collaborative Framework with Extended Mobile Participation for Next Generation Networks. IEEE Access 2016, Volume: 4, Pages: 9129 - 9144, DOI: 10.1109/ACCESS.2016.2602321
* Mariwan Ahmed, Lu Liu, Bo Yuan, Marcello Trovati, James Hardy; Context-Aware Service Discovery and Selection in Decentralized Environments. 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing. IEEE Conference Publications, Pages: 2224 - 2231, DOI: 10.1109/CIT/IUCC/DASC/PICOM.2015.329, Nov 2015
* Lei-lei Shi, Lu Liu, Yan Wu, Liang Jiang, James Hardy; Event Detection and User Interest Discovering in Social Media Data Streams. IEEE Access Year: 2017, Volume: PP, Issue: 99 Pages: 1 - 1, DOI: 10.1109/ACCESS.2017.2675839
* John Panneerselvam, Lu Liu, James Hardy, Nick Antonopoulos; Analysis, Modelling and Characterisation of Zombie Servers in Large-Scale Cloud Datacentres. IEEE Access Year: 2017, Volume: PP, Issue: 99 Pages: 1 - 1, DOI: 10.1109/ACCESS.2017.2725898
* James Hardy, Lu Liu; Available Forward Road Capacity Detection Algorithms to Reduce Urban Traffic Congestion. 10th international conference on cyber, physical and social computing. June 2017. Best Paper Award.

# Introduction

## Background

Road congestion is often cited as a significant and continuing problem. Following on from a very cursory examination it soon becomes apparent that congestion is a very complex and imprecisely defined subject area. The subject becomes even more complex when locations, causes, effects, solutions and impacts are considered. The problems and solutions change depending on route types.

Motorway and inter-urban traffic is generally high volume, high speed with physical direction segregation and minimal “in route” destinations, typically junctions are connected by on and off ramps and grade separations such as underpasses, flyovers and bridges. Destinations can be identified features such as junctions, but can also be unidentified features where vehicles are removed from the flow for example roadside parking or arrival at domestic dwellings.

Urban traffic is typically characterised by a high number of dual direction roadways with lower traffic volumes per roadway. There are a large number of junctions and therefore a large number of interaction points. Speed limits are lower and there are a very high number of destinations. Urban roads can be divided into classifications such as terminal, access and transit each having a different use. For example, terminal may refer to a residential housing area with very low traffic volume, low speeds and traffic control by convention (give way markings) rather than legal instruction (traffic lights)

Traffic flow is studied to make most use of the available capacity. Traffic flow is frequently compared to wave motion, as first credited to Lighthill and Whitman in the 1955 “On kinematic waves. II. A theory of traffic flow on long crowded roads” [1], and to Richards in the 1956 “Shockwave on the highway” [2]. The term “congestion” in this scenario is most likely attributed to high density traffic moving much slower than the allowable and anticipated speed limit. Traffic can also be considered to be particulate, consisting of a large number of individual components each of which may or may not participate in the overall mass flow.

A generally accepted method of reducing congestion is to increase capacity by road widening although there is also research that propose reduced speeds to reduce congestion possibly in an attempt to resolve the “phantom traffic jam”. In urban traffic the term “congestion” is more likely to be applied to times when the traffic is stationary for significant periods of time. The stop-start motion is mainly caused by interaction at conflict points and traffic control measures applied at junctions. While there is some value to considering wave motion, particle motion may be more suited to the investigation of traffic flow in a totally urban environment. The study of mass flows is known as macroscopic, while individual vehicles are known as microscopic. Traffic controls, including pedestrian, vary dependant on road type and use.

A perpetual suggestion is that congestion can be reduced and all transport problems resolved by the wholesale acceptance and use of multi-user shared mass transportation methods. This has been an option for millennia, ranging from shared pack animals, communal living spaces, shared carriages and so on. Given that this is and has always been an option and yet the congestion problem persists with single occupancy vehicles, it seems unlikely that there will be any step change in attitude in the near future.

The current suggestion is that autonomous vehicles with shared ownership and on demand use will make personal vehicle ownership obsolete. This will ensure multiple occupancy of vehicles, reducing traffic volume and instantly resolving congestion. The reality is likely to resemble current car-pooling, taxi-cabs or ride-hailing dominated by private ownership for many years to come.

With no changes to current thinking and no obvious clear alternative to transportation congestion problems, there remains an absolute necessity to improve our management of the problems in the immediate and mid-term future.

## Problem Statement

The problem addressed is urban congestion control with the intention to reduce the negative impacts of congestion by reducing the duration of the congestion period. The project is not to resolve congestion as this may not be practically feasible in any high population density location. Further to this, it is shown that traffic and transportation offer many socio-economic advantages to an area and its inhabitants; such benefits would be negated if traffic were to be reduced. As a consequence, and for clarity, the purpose of this research is to reduce the congestion effects without reducing the amount of traffic or number of journeys undertaken.

Congestion directly affects everyone in the developed world and many people in developing nations. Indirectly it affects the whole planet including uninhabited regions. While time delays due to congestion have been reported since Roman Times, the massive but indirect effects of congestion have only become globally significant since the advent of prime movers which rely on combustion processes to create motive power. It is worth noting that congestion is not limited to vehicular traffic; the subject of congestion in pedestrian traffic (without any vehicles) is widely studied to develop evacuation plans and crowd control processes [3] [4] [5].

Some of the globally significance effects of combustion derived power are:

* the rate of use of non-renewable resources including coal, gas and oil
* release of toxic compounds such as carbon monoxide (CO) and Nitrogen Oxides (NOx)
* release of environmentally damaging compounds including greenhouse gasses (GHG) such as carbon dioxide, nitrous oxide and acid rain components such as sulphur dioxide (SO2)

Locally, the release of heavy particulate matter (soot) is damaging to health and the built environment while the proximity of vehicular traffic will directly result in a higher accident rate and potentially a higher injury rate.

Alternative fuelled vehicles (AFV) are available but this does not resolve the immediate problems for several reasons:

* The power source for pure electric vehicles is likely to originate from fossil fuel power stations. In the UK during 2013 power derivation consisted of 70.4% coal/gas/oil, 17.6% nuclear and 12% renewable sources [6]
* Low uptake. The Society of Motor Manufacturers and Traders (SMMT) figures show 1.38 million new car registrations in Q1 and Q2 of 2015 of which less than 38 thousand were some form of AFV [7]. Conversely, over 97% of vehicles sold were traditional combustion powered.
* The majority of new electric AFV are hybrid electric / internal combustion rather than pure electric (23,500 v 14,000 vehicles in Q1 and Q2 of 2015) [7]. Therefore, over 62% of vehicles were combustion capable.
* Solar cells and batteries have finite lives and need ongoing replacement. Construction methods and materials are hazardous, disposal may be a developing problem.
* Electric vehicles use energy while occupied due to heating, cooling, lighting, entertainment etc. The energy is wasted if the vehicle is serving no useful purpose due to being stationary.

Manual transportation methods such as walking and cycling do not satisfy the needs of developed societies who require bulk deliveries of food and other goods into the city centres. Manual methods are not suited for less able-bodied, longer distance, substantial load, all weather conditions, professional business or events requiring fast response. Public transport systems are limited by destinations and not well suited to lower density suburban or rural residential areas. In urban areas, road based public transport held in congestion may create greater environmental problems; prioritising busses over cars serves to increase the delay for the higher numbers of cars.

There is no simple universal solution. Reducing congestion duration in urban areas has the most impact, is achievable in the shortest timeframe, does not require cultural shift and will continue to be effective in future systems.

Transportation has played a fundamental role in the developing world and will continue to do so. It is inevitable that new or alternative power sources will be developed and utilised in the future to maintain, among other things, transportation

## Aim and Objectives

### Aim

To reduce the duration and negative effects of urban congestion through the development of a novel traffic control method.

### Objectives

Undertake a literature review of existing urban traffic management systems to determine their effectiveness and limitations in measuring and controlling congestion. Periodic congestion is a necessary element in the design of urban road systems. A repeatable and meaningful method to measure congestion is highly desirable for existing control, trending and forward planning and will be a fundamental input to future integrated Intelligent Transportation Systems.

Develop a novel traffic control process which utilises available forward road capacity detection and locality awareness in conjunction with existing backpressure methods as a means to reduce the overall duration of congestion within an area. Urban congestion develops in an area from a single junction, multiple congestion pockets will form bounded by uncongested regions. The boundary locations are both variable and dynamic, the general location changes with time of day and the specific location changes with instantaneous traffic flow. Determining the locally available capacity ahead of the current location will automatically identify the boundary location and allow an even vehicle distribution, reducing traffic density and increasing traffic flow. Identifying the congestion area overall exit boundary will permit the rapid removal of vehicles from the congestion system.

Use formal methods to prove the feasibility and effectiveness of the proposed control algorithm over existing processes and confirm the congestion measurement capability of the proposed controller. By determining the congestion in an area it is possible to influence impending traffic flows by means of reducing the attractiveness of a route in automated guidance systems, automatically diverting traffic by means of upstream controller and variable message systems (VMS) and by preventing traffic from entering the area by gating.

Develop a simulation model of a complex roadway environment and implement a controller as a means to evaluate the potential effectiveness of the proposal. The use of commercially available simulation tools which are widely utilised in the design of real roadway systems will provide a high degree of confidence in the overall performance of the proposed system.

Congestion is a function of population density meanwhile car ownership is considered to be an indicator of affluence; given these two points it can be seen that road congestion can be observed as a positive indicator of local economic and social prosperity. Local politics and economics are generally outside of the scope of this research but their importance and influence must be observed. In order to avoid any unintended consequential damage to the potentially fragile relationships between commerce, economics and population, the solution must not change the number of vehicles attempting to use a given route at a specific time or day. This provides the apparently unsurmountable challenge of reducing the problems of congestion without reducing the numbers of vehicles that are causing the congestion. While this suggests that the total eradication of congestion could be economically and politically damaging, reducing the undesirable consequences and side effects will offer political, economic, social and health benefits. The negative conditions that can be addressed include transit delays, inefficient fuel use, chemical pollution, noise pollution from standing traffic and seismic damage due to low frequency conducted vibration.

An obvious method of reducing congestion is the creation of bypass routes. Besides potentially damaging the local economy, this approach is expensive, takes time to plan, time to implement, causes disruption, is specific to an individual situation and is potentially impossible to reverse. The ideal solution should be universally applicable and immediately effective requiring no changes to road layout or lane use.

Universal applicability implies that the solution will be able to resolve approach congestion from any direction. While not precluded, there should not be a need to specify a primary route which has the beneficial consequence that the system will resolve unexpected congestion caused by nearby accidents or activities. The other major consequence is that there would be no need to tune the signal system for morning and evening tidal flows.

A final objective of the proposed solution is to create a system where adjacent junctions are beneficially interactive but not reliant upon each other. There are two main reasons for this: to avoid failure propagation and to be able to implement the system progressively. From a practical perspective, the ability to implement the solution in stages without major financial penalty, service deterioration, sequential / continuous roadworks and without the need for temporary and costly interim measures would be commercially attractive. The system would offer immediate benefit at modified junctions without the requirement to complete every junction in the area.

## Research Justification

This research is not intended to show any bias for or against traffic control signals, the intention is to identify an improved method of operation.

Traffic congestion is a relatively poorly defined term; it is taken to mean a region and period where travel along a transportation network is substantially impeded by slow moving or stationary traffic. This research is focussed on motor vehicle on urban road networks however the term “traffic” could apply to pedestrian, cycle, rail vehicle, rail passenger etc. and the network may be a road, footpath, railway, flightpath and so on.

Congestion typically occurs when traffic queues are not resolved within a reasonable period. In the language of queueing theory, queues occur when the number of customer arrivals exceeds the service rate of a control point. In urban road networks, intentional control points are typically junctions; unintentional points may be introduced by on street vehicle parking or by incidents and accidents. A junction is an area where vehicles cross paths or flows merge.

As it is not possible for two vehicles to simultaneously occupy the junction conflict area, there will be a need for some form of temporal allocation management system. The allocation system cannot advance the arrival of any vehicle to a period before “t0” and so the only option is to delay one or more vehicles which are contending for the conflict area, therefore traffic control systems always introduce delays.

Junctions are created out of necessity; they are places where vehicles will interact. The interaction will be in the form of contending for the area which is common to multiple routes. It is not possible to create non-interacting routes that simultaneously allow all possible points of origin to connect to all possible destinations.

There are three main conflict management systems in common use: traffic island (roundabout), passive give-way/stop marking and active management (traffic lights).

Traffic islands and passive control allow for opportunist traffic movements based on driver perception, consideration or trust. Opportunist movement can be advantageous when traffic flow is low, inconsistent or intermittently stationary however fast moving continual flows can prevent access from minor entrances as shown in Figure 1.1. The figure shows a condition where continuous traffic from A to C and from C to A prevents the opportunity for any flow from B or from D to C. Traffic islands can become congested due to inconsiderate use which may result from driver belief that there is unfair access. While uncommon, there is a potential for gridlock to arise. Figure 1.2 shows a situation where A to C traffic is blocking D to C traffic which is blocking C to A traffic which is in turn blocking D to A traffic. Traffic Islands are not commonly used in inner city areas mainly because of the amount of space required. Passive junctions, for example crossroads, can suffer from the same problems.

|  |  |
| --- | --- |
| Figure .: Fast Moving Traffic Preventing Access | Figure .: Potential for Gridlocking |

To minimise the overall delay in all directions and to avoid infinite queue formation, the allocation system should allocate temporal space based on direction arrival rates. Only active management can achieve this goal as both road marking and traffic island rely entirely on the autonomous knowledge and actions of each individual vehicle. Hybrid islands, which include active traffic signals during peak periods, are intended to introduce a degree of fair allocation.

Urbanisation provides increased opportunities for employers and employees, it also provides better access to facilities such as medical services, schools and shops. Common start times have developed to be the most practical and commercially advantageous solution; shops do not stagger business hours to open only when there are no customers simply to avoid travel delays; a formal classroom lecture would not be feasible if the lecturer and students were all timetabled to arrive at significantly different times. Social and economic development of towns and cities has led to so called urban-sprawl and the segregation of commercial, industrial and residential areas. As a result of these concepts, there is commonly a need for “bulk transfer” of people at specific times, a period commonly known as the rush hour.

The problems of congestion are not a modern problem. It is reported in various sources that congestion management was an active subject in Ancient Rome [8]. From the reference it can be concluded that congestion was considered a problem at the in two time periods and in two locations: in Rome during the reign of Caesar and in the UK at the time of the article in the 1960s. Other references include 1660 when King Charles II proclaimed that stationary traffic (parked horses or vehicles) should be banned in Westminster and the City of London [9] and in the Stage Carriages Act of 1832 which permits and promotes the use of multi user carriages over single user Hackney Carriages. The arguments of public transport bus services over taxicabs are still important today. As each of these references are taken from times before modern cars were even conceived, it should be clear that congestion is not related to the motor vehicle. More than this, as the references also predate the world’s first, but short-lived, road traffic light in 1868, it can be surmised that urban congestion is not entirely caused by the traffic light.

There are occasional reports of traffic signal failures and of deliberate signal removal. Anecdotal evidence is then provided to justify a suggestion that congestion will be resolved if all traffic signals are removed [10]. There is also any number of headline reports available across the Internet showing where failed signals have led to major delays. The primary purpose of the traffic signal is for safety; to allow conflicting traffic to safely utilise the same road space by temporal segregation, not to enhance traffic flow. It is very apparent that not all traffic signals are ideally located, correctly timed or, on occasion, even necessary; removing some signals may improve traffic flow without compromising safety.

Occupied stationary vehicles waste energy creating unnecessary pollution. Even with the engine off, stored energy may still be used for instrumentation, ventilation, heating, cooling, lighting and entertainment systems; the energy will need to be replenished at some point. The pollution is unnecessary as there is no direct benefit derived from the location of the vehicle. While the concept is clear and well documented for internal combustion engine vehicles, alternative fuel vehicles (AFV) including full electric also waste energy for the same reasons. For AFV, there are also environmental implications of the source of the electrical power, processing and chemicals used during manufacture of the energy storage system and in the ultimate disposal of the storage system. Batteries generally have finite lives, needless discharge during queueing and consequential recharge will have an impact on the battery replacement interval.

Autonomous vehicles by definition are not aware of the general state of other traffic and cannot make universally beneficial decisions. Connected vehicles will similarly not be capable of making universally beneficial priority decisions without knowledge of the surrounding traffic situation including road capacity and current queue status. This information, including own vehicle priority, would need to be sourced from an independent traffic management system.

In summary: Successful urban living necessitates transportation. Diversity of origins and destinations requiring near simultaneous access leads to conflict between traffic flows. To allow safe passage through contended areas, a prioritisation or allocation system is required. To ensure a reasonable degree of fair access the traffic management system must be both active and independent. The management system will introduce delays. Delays will be in the form of queues which will be observed as congestion if not cleared in a reasonable period. This is equally applicable to all forms of transportation. Powered vehicles, of all types, are responsible for environmental damage; vehicles in queues create environmental damage without any tangible benefit.

Active traffic control systems will be a feature of urban transport systems for the foreseeable future. Research towards reducing urban congestion by improving active traffic management is totally justified by the need to limit environmental damage, improve economic efficiency and improve urban mobility. The specific problem of exit blocking as highlighted and the solution presented in this research will remain equally applicable for future traffic management systems.

## Research Novelty

This research specifically identifies a method of managing road congestion as opposed to managing traffic flow in an urban environment. This is achieved through the definition of a new algorithm that creates direct prioritisation based on exit detection to determine the current ability of a junction to fulfil a demand and downstream communication to ensure the ability is available in the immediate future. The novelty can be observed by considering the action of existing traffic control signals which are a method of safely providing a fair allocation of contended sections of roadway. Fair allocation is achieved by increasing the priority of a specific flow direction and consequently reducing the priority of the conflicting flows. The conflicting flow priority can only be theoretically reduced until it becomes zero, at which point the demands for increased priority can no longer be satisfied and active control is not possible. The measurement of demand is determined by detecting approaching traffic and there is an un-confirmable assumption that the flow directed by the signal can always be satisfied because the exit route is clear. By definition therefore, existing traffic signals can only actively manage moving traffic and therefore cannot identify congestion until it has occurred whereupon they have no capability to control or dissipate congestion.

The downstream communication method is novel in respect of being able to form a loosely coupled community with another signalised junction without additional planning or programming. This provides the ability for junctions to be integrated, added or removed without significant alteration to existing traffic management deployments which is in stark contrast to the complex planning required for existing UTC systems. The communication and ability to rapidly loose couple traffic signals could be used to integrate and influence “temporary traffic lights” to minimise the insertion delays caused by roadworks.

This research considers congestion across an entire road network with all routes and destinations considered equal. Multiple signal controlled junctions are simulated and the traffic flow is considered from every direction and using every path simultaneously. This is in direct contrast to many research and production systems that consider either a single junction, a single direction through a junction or a single flow direction / path through a network.

The research proposes an algorithm based on a easily obtained additional information source not used in existing systems. The algorithm minimises the traffic flow time lost by existing methods without compromising safety.

The algorithm is specifically intended to avoid, manage and resolve congestion rather than the traditional approach of attempting to manage traffic flow.

The research defines a self-adjusting, self-tuning control system that adapts to actual current flow restriction patterns without time/day based predictive algorithms. The major advantages to this methodology are no timing alterations are required to accommodate tidal flow directions and that the entire road system will respond to unpredicted, unanticipated and unexpected flow modifying events such as accidents and social gatherings.

One of the findings of this research is that it is not possible to accurately predict traffic conditions and therefore a system that reacts to actual measured conditions is necessary to effectively manage traffic flow. The inability to predict traffic conditions is based on Chaos Theory and is supported by the trend adjustment algorithms used in commercial traffic control systems e.g. SCOOT and SCATS.

The outcome of this research is a novel control system algorithm which has the ability to reduce the level of congestion in an urban area by employing the following unique features, attributes and capabilities:

* Confirming the ability of a flow to commence in advance of a flow indication. Existing systems assume that a flow will occur when a green signal is shown.
* Active calling of red indication if an exit becomes blocked during a flow period. Existing systems only actively call green indications; red indications are a consequence of a green called from an opposing flow.
* Minimise transition losses under certain circumstances by avoiding minimum green, amber and “all red” time periods.
* Ability to form on demand, ad-hoc signalling networks with elastic boundaries to promote congestion clearance. Ability to determine the current congestion boundary. Existing controls are either isolated, form part of a pre-determined network or are limited by the network that can be formed.
* Highly dynamic response to actual traffic conditions across a wide area in real time. Existing advanced urban traffic control systems apply a timing solution based on historic flow data and predicted outcomes; to avoid instability the timing decision is made over several cycles and the amount of change is limited. The solution is therefore highly reactive and has limited capability to resolve situations that are highly dynamic or unusual.
* Ability to simulate opportunist movement similar to traffic islands and formally impose simulated courtesy observable during certain junction signal failures.

Existing control systems attempt to share access to the conflict area (controlled junction) in a fair way, based on traffic arrivals. The sections of the local road network affected by congestion vary, with pockets of heavy congestion bounded by light traffic. Vehicle Actuated (VA) controls are effective at managing the traffic until the traffic volume exceeds the junction capacity at which point traffic exceeds the detection limits and the controller falls back to a fixed timing system which may or may not be optimised for the prevailing situation. Congestion is then present until it “naturally” subsides and active VA control resumes. Flow through the junction is achieved by allocation of a suitable length of green time, which may be called, extended or curtailed as required. In the UK, an impeding change to a green signal is indicated to vehicles by a fixed time period of combined red and amber (yellow). Green periods must exceed a minimum period, the end of a green period is indicated by a fixed time period of individual amber. There may be a period where all directions are shown red indications to allow traffic to clear the junction before allowing a conflicting flow to begin. It is assumed that traffic will move when shown a green indication.

This project will develop a new control method that uses traffic arrival as per existing systems but considers the state of the junction exit to assess the available forward road capacity to enhance measurement and control. By considering the exit state, a signal change from an existing red can be avoided if the exit is already congested preventing an unnecessary flow loss. Also, in the event that an exit becomes congested during an existing green, an early change to red can be commanded. In this circumstance, traffic must already be stationary and therefore the minimum green time and yellow period have no useful purpose. The maximum amount of time is therefore given to allocation of green time to flows which are achievable. The demanded red time is of no benefit to the direction to which it is applied however it is not detrimental either, as the flow is physically blocked and could not occur regardless of indication. Conversely, the additional green time allocated to a conflicting direction is beneficial to that flow and so overall congestion is reduced and the efficiency of the entire junction is improved. Early clearance of one direction will reduce the need for signal changes, and hence transition losses, during later cycles leading to a reduction in the overall congestion period. In a separate but related activity, the reason for the exit being blocked is actively addressed by message to the next signal downstream of the blocked exit. The message is a request to service the blocked route as a priority. If the downstream signal is also experiencing problems, it will make another request to downstream neighbours and so on until the current dynamic edge of the congestion is identified. Releasing traffic out of the congested zone as a priority will reduce the congestion area and resolve the problem in the least possible time. The period of time that the traffic control sequence is dependent on the junction exit state will directly and accurately indicate the level of congestion along that route, information that could be used by traffic planners or traffic guidance systems.

## Methodology

### Research Methodology

Internet, library and local council research has been undertaken to identify operational and legal requirements for traffic control systems. Existing University and Derby City Council connections were included where possible and complimentary new links have been forged with the City Council Traffic Management group.

Academic and transportation biased research continues to be studied to understand in depth the problems and benefits of traffic congestion along with potential solutions and impacts. The research methods and solutions identified in other works have been studied to their similarity and applicability to this research. Congestion metrics including “lost time” and “wasted fuel” have been investigated along with other methods for congestion measurement.

Mathematical methods as used to define and analyse traffic flows and congestion causes have been researched. Assistance from and collaboration with members of the University of Derby Department of Mathematics has been invaluable in understanding the role and application of Game Theory, Queueing Theory, Cellular Automata and Process Engineering in the traffic management process.

Human response and processing time is considered to be vital in understanding rates of queue creation and clearance time. Researchers from the University of Derby College of Life and Natural Sciences and College of Education have been contacted as sources of information regarding human response and reaction times. The response and processing time for automated vehicles and driver assistance automation is likely to be much faster than for human beings however it is still a finite value. The triggers, delays and reaction times are important for predicting the suitability of future Intelligent Transportation Systems.

Personal membership (and subsequent enrolment of the University of Derby as a corporate member) of ITS-UK has proven highly beneficial in providing cutting edge research materials, seminars and access to influential Government bodies, Local Authorities, Universities, researchers and research establishments and implementation businesses. ITS-UK is a not-for-profit association promoting the development of Intelligent Transport Systems (ITS) in both public and private sectors. Objectives of the various working groups include improving economic efficiency and transport safety while minimising negative environmental effects through the development of the ITS market. Group meetings of the “Local Authority and Urban Interest Group” and the “Connected Vehicles Interest Group” seminars have been attended and a personal contact portfolio is being continually developed.

Information has been exchanged during attendance and participation in group meetings of IMPart (Intelligent Mobility Partnership), a consortium of researchers from the universities of Loughborough, De Montford, Nottingham Trent and Coventry alongside the Transport Systems Catapult based in Milton Keynes.

Knowledge and experience has been gained from seminars, meetings and discussions with individual members of the Spanish and UK technical and research groups from Transport Simulation Systems, the producers of the Aimsun traffic simulation package.

### Project Methodology

The subject field is very large and it was necessary to accurately define the boundaries of the research and its proposed outcomes.

The problems and benefits attributed to congestion were considered to ensure that the project was both viable and beneficial. Due consideration was given to the idea of what could be improved as well as what should be improved. A method and means to quantify the existing situation and therefore provide a comparative measurement of any improvement was determined.

Methods used to control, measure and manage traffic flow were investigated including formal mathematical procedures, modelling and simulation systems. Benefits and limitations were considered including the ability to practically assess the outcome of the proposed system.

To enable simulation to be used as a viable method of obtaining results, a test scenario was created including a hypothetical layout and traffic flow. The choice of simulator, scenario and any assumptions are fully justified. The traffic light control system, written in Python2, was created outside of the simulator and the actions applied using the simulator API. This approach ensured that the control system was totally defined and understood. Three control systems were defined: fixed timing, vehicle actuated and the novel approach of available forward road capacity. Each control system was applied to simulation and the results obtained. The results were examined and conclusions were drawn which allowed an iterative development process to be undertaken for the simulation as well as for the control system.

At the end of the project, overall conclusions are presented regarding the success of the project and suggestions are made regarding practical implementation and future development opportunities. The near-term, mid-term and long-term benefits of the system are identified. The relevance of the project to potential transportation developments such as connected and autonomous vehicles is also considered.

## Purpose of Research

This thesis is primarily concerned with managing traffic congestion in urban road systems. The project rapidly converges towards the traffic light for several reasons, namely that it is the most common form of active urban traffic control, provides a sequence of legally enforceable transitions and it can be the cause of congestion incidentally or by design.

The project examines the causes of congestion in urban environments and attempts to address the negative aspects while recognising the inevitability of congestion in urbanised areas. The project proposes a new control method to address the clearance of pre-existing congestion as well as reducing congestion build up.

Existing actuated controls react to approaching vehicles. The green time response may be directly proportion to the current traffic arrival rate, to the average queue at the end of the green time in a number of previous cycles and/or the rate of change of the reminder queue. Dependant on the system employed, overall control is either centralised across an urban region or isolated to a single junction area. Detection is commonly a loop detector embedded in the road surface though other detection methods are available including optical recognition, RADAR and LIDAR. When the traffic volume reaches the saturation level at a junction, queues form which may rapidly extend beyond the detection range. As the queues at this stage are effectively infinite, the anticipated traffic control response is to increase the green time duration to a maximum value in each direction. At this stage the controller will effectively be operating as a fixed time system.

This project develops a method of congestion detection and an algorithm to dynamically resolve the problem. The controller is built on pre-existing control technology utilising the benefits of well-known systems but addresses the known problems by the addition of an additional exit state detector.

The main motivation for this project is the need to reduce the detrimental environmental impact of transportation. To achieve this it is necessary to understand why traffic congestion occurs and whether it can be managed more effectively than current solutions permit. This is not a trivial task, congestion has been a long term problem which has proven elusive to resolve.

For this project, the initial problem was contrived from an observation of congestion level and timing in a specific area leaving the city of Derby, England by a major route towards th North. The route is shown in Figure 1.3 (taken from Openstreetmap.org) and includes Eastgate and the Inner Ring Road (A601), A6 King Street, Quaker Way, Lodge Lane junction, Five Lamps junction and exiting towards Duffield / Matlock. The route provides a major corridor from mainly residential areas to the North of the city (Allestree, Darley Abbey, Duffield, Belper, etc.) towards the City and onwards to commercial and industrial areas to the South (Allenton, Sinfin) and South-East (Pride Park) and also West towards Nottingham and the Southbound Motorway network. The location of the residential centres and mass employment areas ensures that traffic is highly tidal in the area and results in significant congestion northbound during the evening rush-hour and southbound in morning, albeit by a slightly different route. Queues are minimal other than at the peak times.

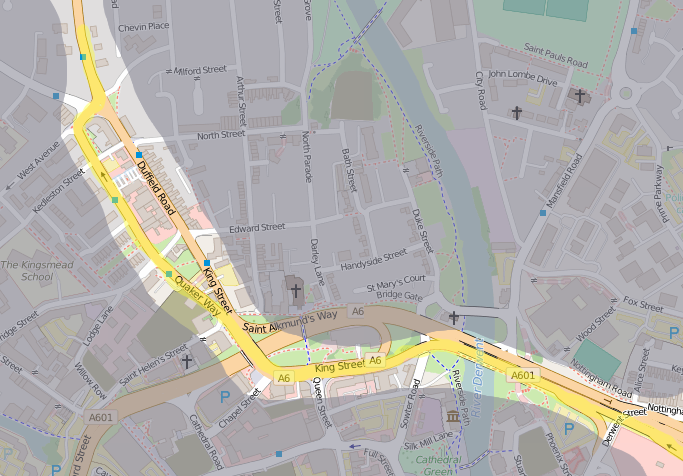


Figure .3: “Derby Junctions” Road Map from Openstreetmap.org

The observations included identifying that traffic lights and pedestrian crossings did not appear to work in conjunction to resolve the congestion. There were periods when the exit from one junction was blocked by standing traffic which was the result of a traffic light further along the route. Referring to the map shown in Figure 1.3, relatively minor traffic flow entering King Street from Queen Street appeared to have priority over the main King Street traffic. There is sporadic but significant traffic flow from both directions on Lodge Lane into Garden Street. Significant traffic travels north along Garden Street and makes a 180° turn to travel south along Duffield Road; between 3 and 5 vehicles following this route can block access to the major route to the north along Duffield Road. Traffic will be queueing along the right hand lane of King Street, Quaker Way and Garden Street during the evening rush-hour, a number of drivers will use the left hand lane until near the Duffield Road junction and then make an impertinent move to the right hand lane. While the move is not illegal, it clearly causes a degree of frustration at times with drivers physically challenging for road position and displaying disapproval with visual and verbal communication. Vehicles that exit to Duffield Road heading north will leave the congested region and are unlikely to enter another queue for over 1.5 miles; this seems to apply at all times of the day. Traffic heading south from the Garden Street / Duffield Road junction *may* leave the congested area but then enter another congestion region either at Lodge Lane crossing Garden Street or, more significantly, joining the Inner Ring Road which will commonly queue along Eastgate towards Nottingham. The queue will clear at the junction with the A52.

According to Google Maps, in 2016 driving the 1 mile between The Underpass and Duffield Road junction was predicted to take 4 minutes on a weekday at 1pm, an average of 15mph. The same journey was predicted to take 7 minutes at 5:30pm (average 8.5 mph) and 4 minutes again at 6:30pm. A limited amount of practical testing was undertaken in 2016 and showed that in a sample of 5 journeys at each time, the journey took 2.25 – 3.5 minutes at 1pm, 10.5 minutes at 5:30pm and ~7.5 minutes at 6:30pm. Except for 1pm, the times were consistent to within a few (<5) percent. In each case, the maximum speed limit was observed, there were no unnecessary lane changes undertaken, the correct lane was selected at the earliest reasonable point. The route is approximately 50% in a 40mph zone, 50% in a 30mph zone and passes through 4 sets of traffic lights. Google had updated their estimated journey time for 5:30 in October 2018 to a range of 3-10 minutes however the times for 1pm and 6:30pm were unchanged.

In summary, the variation and inconsistency of these results show that the traffic flows are complex, temporal and tidal.

There have been signal and layout changes to the roads in recent years intended to improve access and reduce congestion, the results of which are open to opinion and interpretation. In defence of the planners, there are a large number of historic and listed buildings in the area that limit the space and opportunity available for a large scale redesign involving route changes and road widening. Given that flow improvements cannot be addressed by roadway alteration highlights the need for a different method of flow control.

Table .1: Google Prediction Compared to Actual Journey Time and Speed

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Time | Google Prediction | | Actual Journey | |
| Journey Time (minutes) | Ave. speed  (mph) | Journey Time (minutes) | Ave. speed  (mph) |
| 13:00 | 4 | 15 | 2.25 – 3.5 | 17 - 26 |
| 17:30 | 7 | 8.57 | 10.5 | 5.7 |
| (17:30, Oct 2018) | 3 – 10 | 6 - 20 |  |  |
| 18:30 | 4 | 15 | 7.5 | 8 |

Despite being the basis of the study, the named Derby junctions were not used to develop the theoretical solution. After due consideration, it was decided to define a grid network similar to the Manhattan / Milton Keynes road systems for the following reasons:

* The Derby junctions route is not intended to be a damning indictment of the City of Derby which should be resolved above all others, it is intended to serve as a practical illustration of the problem under consideration.
* Most importantly, the local problems at any specific junction are an engineering problem. It is possible to completely miss the point of this research by discussing the relative merits of grade separations, alterations to lanes, introduction of traffic enforcement cameras and so on for a specific area.
* The theoretical solution proposed by this thesis is intended to be universally applicable and not restricted to solving the problems associated with a single junction. The intention is to determine a new general strategy for controlling traffic under the very specific conditions of temporal congestion at sequential active traffic control points.
* It is relatively easy to obtain the detailed features of the road from a source such as openstreetmap.org and convert the information to a form used in the traffic simulator, Aimsun. Aimsun provides methods to automate this. Openstreetmap is highly detailed and visually complete however a significant amount of time may be necessary to correct alignments and features. The Aimsun academic licence is limited by the length of road and number of nodes (joins between features such as two roads or a road and a junction), the import area contains many features that are not relevant and easily exceeds the licence restriction without substantive benefit. Conversely, some of the side streets provide practical alternative routes that may be highly significant. Traffic lights and detectors need to be added manually, the sequencing of the signals is unknown. The result is a time consuming imperfect representation which may not be capable of generating valid results.
* A valid simulation of the roadway requires real traffic data, flow rates and signal timings. Some of the information, gained from in-road detectors, is available from the Local Authority. Information along the “Derby junctions” route can be obtained from the SCOOT detectors, however the nature of SCOOT means that individual turn information is not available. In summary, the information may be incomplete, time consuming to obtain and, ultimately, of minimal value to the research.
* A theoretical network can be defined to have multiple junctions and alternate routes. For the research project this offers many benefits including a high level of interaction between each junction and the lack of a preconceived best route. The theoretical traffic volume can be adapted and maintained to ensure both over and under loading of the road network to test the capability of the solution.

This research develops novel control algorithms to actively manage congestion through road space allocation rather than attempting to manage approach flows. A direct consequence is the ability to capture congestion metrics for use in related systems. One of the goals of the research is to identify a practical system that can be implemented in the immediate term using existing traffic control hardware with minimal change. The control methodology will reduce congestion effects for all modes of urban road transport and will be equally applicable to existing vehicles and future connected and automated vehicles without significant additional road infrastructure disturbance and expense.

## Scope of Work

This work is limited to the effective active control of traffic congestion in urban environments in the present day and near to mid-term future. The primary focus is toward existing vehicles and technologies. The work shows the advantages that are immediately available and briefly considers the related subject of advances in vehicle technologies. There is minimal reference to the problems of inter-urban traffic congestion which typically include different control methods and causes.

While not definitive, the following explanatory notes are included for clarification:

Urban traffic systems are lower speed networks, typically with speed limits in the range 20 – 40mph, and occasionally as high as 50mph. Physical lane segregation may exist, but bi-directional single carriageways with white line lane division is normal. Traffic volumes have a very wide hourly average range over cyclic days, cyclic weeks and cyclic seasons. Hourly volumes are influenced by local events including working times for local industries and schools, presence of retail areas (as opposed to isolated retail outlets) and mass attendance venues such as for sporting and entertainment events. A large variety of vehicle classes are expected including bicycles, invalid carriages, low speed and high speed motor cycles, small and large private cars, public transport single and double-deck busses, trams and commercial vehicles ranging from taxis and small vans to six-axle, 44 tonne articulated lorries [11]. The probability of interaction with pedestrians and cyclists of all ages and abilities is very high as is the probability of interaction with small animals for example cats, dogs, urban foxes etc. The likelihood of interaction with large animals, for example horses, is low but feasible. Road marking and street lighting is normally good, though there is an apparent trend towards reducing street lighting as a means of cost reduction to local authorities. There are typically a large number of traffic junctions and traffic control method is most commonly stop/give way and actuated traffic lights. A range of pedestrian crossing methods will be employed including bridges, subways, co-operative “zebra” crossings and multiple versions of active signal controlled systems. Within the urban area there will be different types of roads with different purposes for example arterial routes, access routes and residential streets. Parking is not permitted at certain times on Urban Clearways which are specifically identified but artificial road and visibility restrictions caused by parked cars are commonplace in most other urban areas.

Inter-urban road systems are usually multi-lane, high speed, high volume dual carriageways and have good distance visibility. Dual carriageway means that there is physical segregation between traffic directions using e.g. “crash barriers”. The roadway is normally segregated from urban areas by distance however additional protection to prevent accidental incursion of people and animals is usually required. Inter-urban roads can be divided into motorway and non-motorway major routes. Motorway routes are very high volume, normally very high speed 70mph with multiple lanes. Location, access and legal restrictions mean that traffic is normally limited to high speed motorcycle, private car and commercial vehicle with a high incidence of large articulated lorries. Junctions are conventionally located off the main carriageway and therefore conflicting traffic flows rarely interact within m-way regulations. Interaction with pedestrians and animals of all sizes are extremely rare. Motorways are normally constructed to pass near major towns and cities with high speed major routes to connect to the towns and cities. Major routes are typically A-class roads and commonly dual-carriageway. Typically the speed limits are either 60 or 70 mph. While A-class dual carriageways share some attributes with M-class roads, there are generally no legal restrictions on low speed motorcycles, bicycles and even pedestrians. There are normally an increased number of junctions and the junctions are likely to be roundabout junctions which are in line with, and interrupt, the main traffic flow. The probability of interaction with pedestrians and small animals is greatly increased as the roads become more accessible and are closer to residential areas. The roads are normally “Clearway” routes meaning that there should be no parked vehicles to affect flow or visibility.

Rural roads are typically national speed limit single carriageways (60mph) with indeterminate visibility and carry low volumes of traffic. Give way junctions are the most likely method of traffic control with some traffic lights used only at the most major junctions. No vehicle classes are prohibited by legal means. The probability of interaction with bicycles, slow moving agricultural vehicles and large animals, e.g. sheep, cattle, deer, etc is much higher than for other road classes. The combination of effects can make rural roads statistically the most dangerous in terms of fatalities.

Upstream refers to the direction from which traffic will, under normal conditions, arrive at a specified location. A junction will have more than one “upstream” direction.

Downstream refers to the direction that traffic leaving a point will travel towards under normal circumstances. A junction will have one or more “downstream” directions.

One or more vehicles making, or intending to make, a transition from an originating location to a destination location is known as a flow. The origin and destination may be arbitrary or final. A flow through a junction will have only one “upstream” and one “downstream” at any time. A junction will always have more than one potential flow and it is normal for more than one flow to be active at any time.

A typical simple junction is referenced throughout this work. The junction is shown in Figure 1.4 and clarifies the direction references.

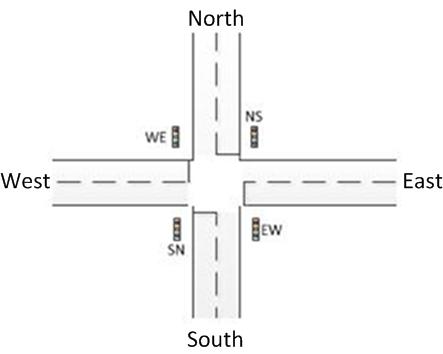


Figure .4: Junction Directions

A conflicting flow is a flow which cannot safely occur simultaneously with another flow under consideration. For example, at the simple crossroads, a flow from North to South is in conflict with a flow from East to West.

An opposing flow can normally coexist on the same roadway when the roadway is subject to some conventional, physical or logical division. For example, a flow from North to South is opposing a flow from South to North and both can simultaneously utilise the same “two-way street”.

The contended area is the area within a junction which is used by conflicting flows at different periods of time. As an alternative definition, it is the area which cannot be utilised by two flows simultaneously. There may be more than one contended area in a junction. The action and physical attributes of junctions are such that the per flow upstream and downstream capacities are significantly higher than the contended area average capacity.

Isolated junctions are those which are not affected by, and do not interact with, the flow control systems of other junctions. This is typically the result of physical geographic spacing. If the spacing between junctions is reduced, there is an increased probability that the traffic flow at one junction will have an effect on the flow of another nearby junction. A junction that interacts with another is a non-isolated junction.

Traffic flow is considered to be below capacity when the roadway is underutilised and has significantly more capacity than is being used. No vehicles should be impacting the flow of any other vehicles. The vehicular exit capacity for the roadway is higher than the arrival rate, therefore queues do not form.

Traffic flow is deemed to be at (or very near) capacity when the roadway is busy but still flowing. The movement of each vehicle is dependent on other vehicles in the roadway. Exit capacity and arrival rate are similar values, queues form and disperse, the average queue length remains stable at zero.

When traffic flow is above capacity the roadway is congested. While it is not physically possible for the number of vehicles to exceed the available road space, the number of vehicles attempting to use the roadway exceeds the free flow capacity meaning vehicles are no longer able to move freely. No following vehicle can advance until the vehicle in front has progressed. Congestion begins at a specific physical point and grows back upstream from that point; congestion will be caused by a physical event such as a road narrowing or non-uniform / asynchronous vehicle movement. Average queue length is above zero.

In traffic signal terminology the word split refers to the amount of green time as a percentage of the entire signal cycle.

## Thesis Structure

The remainder of the thesis is organised as follows:

Chapter 2 is the Literature Review where the works of other authors in the field are discussed, components of existing traffic management systems are explained and formal mathematical tools are identified.

Chapter 3 defines the problem to be addressed in terms of congestion history, effects and identifies the traffic light as one of the major modern day causes.

Chapter 4 discusses the logical components of traffic systems including the process of traffic modelling. The section considers the detection of traffic, efficiency of transportation methods, traffic flow, types of traffic and the impact of diverging and converging traffic flows.

Chapter 5 explains the proposed solution in detail including the general concept, perceived benefit and the changes to existing systems that would be required in order to create a practical working solution. The section also states the case for the proposed solution to be a fundamental part of any future congestion control process including driverless vehicles. The algorithm design detail is included in the chapter.

Chapter 6 consists of the simulation definition, simulations and results obtained throughout the system design and development. Comparisons are given between the capability of traffic lights based on Fixed Timing, Vehicle Actuation and Available Forward Road Capacity (AFRC) enhanced control, the subject of this thesis.

Chapter 7 provides a list of the conclusions of the work including the technical limitations of the proposed system. There is a discussion regarding potential future transport changes that may affect congestion levels and finally some suggested research that may be undertaken following on from this thesis.

# Literature Review

## Introduction

Traffic control signals have a single primary objective which is to provide a reliable, cost effective method of improving road safety at junctions or other contended sections of a roadway. The safety aspect is achieved by temporal segregation; permitting only one of a conflicting set of flows to use the contended space at any time instance, denying all flows to ensure clearance of the contended space and then passing priority to another flow from the conflicting set. Reliability and cost are managed by automating to ensure continual repetition of the task and to remove the ongoing cost of employing a person.

A secondary function of traffic signals is to provide a statistically fair allocation of time to each of the conflicting flows. Determining the metrics of fair allocation is a matter for discussion. Examples of metrics can be absolute number of vehicles, absolute length of queue, capacity of queue as a percentage of anticipated flow rate, waiting time for the lead, middle or last vehicle, average waiting time for all vehicles and so on. The metrics may also be compounds of elements and may be presented as a single value, sequence of values or a probability distribution. There is significant interest in the statistical analysis of traffic flow in order to identify theoretical improvement caused by the application of alternative timings, methods, metrics or measurements. In order to test the relative outcomes of the mathematical analysis certain assumptions have to be made to ensure full compliance with the algorithm under test and that no other external influences could independently modify the result. The assumptions mean that the vehicle approach rate exceeds the service rate and that there is nothing downstream of the junction. This is the actual definition of an isolated junction that is busy but not saturated; it is not an urban junction with downstream queues restricting outflow and it is mathematically unable to resolve the input queue meaning that it is causing upstream congestion.

The existing traffic signal was not intended to manage, avoid or resolve congestion even though this is a common expectation. Traffic control systems manage traffic flows based on traffic arrivals, prediction and flow data collected from manual traffic counts. In order to manage congestion there must be some form of stationary traffic detection. For existing systems, an increase of flow capacity in one direction is achieved by reducing flow capacity in a conflicting direction. There is a maximum time that a stop indication can be shown before there is a forced transition, similarly there is a minimum time that a signal must show either green or single amber before transition to the next phase. Each transition introduces delays and includes a period of total loss where no vehicles are in the contended area. Existing signals are only equipped with approach detection and, as a result, have limited ability to detect a restricted exit flow capacity during expected traffic flow periods and no ability to detect impending restricted flow during expected stationary periods. A traffic signal phase can interpret approach detection to continually demand higher priority while traffic is stationary or can prioritise a transition to a phase where no flow is possible because the exit is blocked by traffic tailing back from the downstream junction; either case represents a total loss of flow time. The total loss time is added to the journey time for all vehicles in all directions whereas avoiding a total loss in one direction could reduce the total combined journey time for all vehicles by up to 50%, reducing the duration of congestion by the same amount and benefitting all vehicles.

## Related Works

Pandit, Ghosal, Zhang and Chuah present a paper that is indicative of papers that consider co-operative and synchronised flows through multiple control signals, referring to “platoons” [12]. An observed limitation of all of the methods reviewed is that they address traffic flow through ingress rather than congestion by egress. Schemes typically involve attempting to maximise flow in uncongested junctions, maximise capacity at congested junctions or reduce traffic access into an area through “gating”. Solutions generally aim to reduce waiting times by maximising traffic flow at a junction rather than reducing congestion causes. It is common in junction controller research to considerably simplify the models by assuming the junction is isolated [12] or busy but not saturated [13]. Signal timing is invariably modified based on the detection of traffic approaching a junction, this is the case with all known existing systems such as SCOOT, SCATS and RHODES but also with research systems [14]. New technologies such as Wireless Vehicle Ad-hoc NETworks (VANETs) emulate arrival sensors using in vehicle sensors to predict when a vehicle will arrive at a stop line [12]. A paper by Gregoire et al. discusses capacity aware back-pressure control in relation to the problem of unfair service allocation caused by the detection of infinite length queues on junction approaches with unequal lengths [15]. Qi and Zhou proposed an emergency control sequence to divert traffic from an incident located at a downstream junction [16].

A paper from Manolis et al. [17] contrasts the benefits of centralised and decentralised signal controls. The paper suggests that centralised control is not always beneficial and discusses the concept of spillback blocking. There is experimental simulation conducted with the aim of proving the improvements that are possible. The paper presents very clear arguments however it does not resolve all of the points raised in this thesis. Of primary note is that the subject of spillback blocking is addressed by estimating the time / volume of traffic required to cause a problem and then reducing the available green time for this direction as a means of prevention. This is a normal approach however it fails to acknowledge that there may be hidden entrance or exit points in any path; places where vehicles may begin or end their journey or may temporarily stop, effectively ending a journey in one traffic light timing cycle and beginning a new journey in a subsequent cycle. The approach makes a very clear assumption regarding driver and vehicle response, which cannot be relied upon. Finally, if the spillback volume assumption is incorrect, it will lead to lost time due to over-frequent traffic light changes, inefficient road use due to consistently allowing unused gaps in the traffic flow to form or cause / exacerbate spillback blocking of upstream junctions regardless of the upstream controller action. It may be clear from this sentence that spillback blocking is frequently an avalanche effect that rapidly transitions back upstream from an initial point. The next observation of the paper is that the simulated road network identifies very specific entrance and exit points meaning that there area is purely transitory with no origins or destinations within the network (including totally internal journeys). This leads to the final observation that the simulation may be entirely tidal and not consider minor traffic flows that exist in opposition to the assumed flow.

Parmar, Trivedi and Stevanovic published a paper [18] to dynamically manage signal phases to address traffic congestion. The paper observes the issues of congestion and proposes a resolution based on identifying every vehicle in the region, determining the make, model, physical size and fuel type. The system relies on every vehicle being equipped with a uniquely addressed GPS location device. It also relies on every vehicle and signal being connected to the public internet which it uses as a data transport system to connect to databases and a compute system that supplies the vehicle information, determines vehicle travel history, calculates road occupancy, signal phasing and alternate routing. The paper has a small review of an early AFRC paper [19] and claims that AFRC is complex and that data acquisition is a challenging hardware issue. The authors do not consider the later AFRC paper [20] and have clearly failed to understand the AFRC concept. They imply that there is some network of additional controllers required for AFRC. The paper fails to recognise the importance and influence of the driver, assumes that all vehicles will be connected and never experience failure and that no vehicle will be fraudulently registered (not appearing in a database or simultaneously appearing in more than one physical location). It also fails to recognise the logical rather than physical aspects of queueing i.e. that the boundaries of a congestion region are variable over time and at different times.

The literature review has not identified traffic light controls that include a detector and method to confirm an exit is clear before changing to a go signal which is proposed as a novel aspect of this work. Whilst effective under normal conditions, detecting or predicting only traffic ingress at a junction under periods of congestion could cause existing control algorithms to allocate longer go signal durations to routes which are already blocked. This is counterproductive and potentially creates or increases congestion in cross routes.

Methods of describing and quantifying the main issues related to traffic congestion have been identified [21]. Metrics are required to enable direct comparison between model, simulation and physical measurement. Classical mathematical modelling uses total delay time, number of stops, duration of stops or average journey speed to measure congestion. There are also psychological, interpreted and relative aspects to congestion measurement.

Alternative methods of traffic control have been investigated. “Shared Spaces” schemes aim to equalise the implied priority of every transport class (motor vehicle, cyclist, pedestrian, etc.). The objective is met partially by the removal of traffic light controls at junctions and pedestrian crossings. Amongst others, the shared spaces scheme in Poynton, Cheshire has been hailed as a success by some [22] however this is not consistent with a more general report [23] which states that the general level of safety is reduced. Reports presented at Poynton Town Council meetings suggest that the harmonious interaction between pedestrians and vehicle drivers may have been short lived [24]. The Poynton scheme removed traffic lights at a junction and there are claims and counter-claims that traffic flow was increased as a result. One of the main supporters of the Poynton scheme is highly biased by a personal campaign aimed at the removal of all traffic lights [25]. There are ongoing plans to construct a bypass / relief road around the town to reduce traffic.

It is widely implied that the problems of congestion will be resolved by developing technology. It is also implied that new technology will remove the need for traffic lights. There are numerous topic areas related to connected and autonomous vehicles (CAV) each of which are entire research projects. The fundamental use case and benefits of connected and autonomous vehicles is still undefined with many unsubstantiated claims and counter-claims. It is very unclear how CAV will actually reduce congestion.

## Traffic Light Control Systems

### Outline

The majority of current traffic lights use information gathered by detectors to determine traffic arrivals. The knowledge of prevailing traffic conditions is used as a method of improving junction utilisation and reducing delays. Traffic light systems are programmed to change state according to various information sources. Timing can be varied to predict requirements based on historic junction information and special event circumstances or in direct response to detected traffic. Hamilton [26] explains the different actuation methodologies in terms of developmental progression covering 1868 to present day and describes various production systems including timed, VA, Urban Traffic Control (UTC) (SCOOT, SCATS, UTOPIA, RHODES and MOTION) as well as Intelligent Transport Systems (ITS) and future generation Co-operative Vehicle-Infrastructure Systems (CVIS). Each control scheme has a different method or traffic bias, for example the UTOPIA system offers “Selective and absolute priority to Public Transport vehicles” according to a manufacturer presentation [27].

Traffic detection includes class of vehicle, approach speed, traffic volume and traffic flow. Common detection methods can include magnetic, inductive loop, video (visible and infra-red), laser and radar [28].

Information based systems can be categorised in many different ways. The system can be classified as actuated or adaptive, isolated or coordinated and the control system may be centralised or distributed. Information based systems are not precluded from functioning as timed systems, the choice can be predetermined to a schedule or occur as a result of current conditions.

### Sequencing

The simplest sequence of lights for a single crossroad junction shown in Figure 2.1 is given by . This shows only the essential indications of stop (red) and go (green). NS refers to a signal controlling flow originating from the North and travelling towards South, EW refers to flow from East to West, SN refers to flow from South to North and WE indicates flow from West to East.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Figure 2.1: Simple Crossroad Junction with Signal Directions | Table 2.1: Essential / Minimal Light Sequence   |  |  |  | | --- | --- | --- | | Stage | NS (SN) light | EW (WE) light | | 1 | Red | Green | | 2 | Green | Red | |

For increased safety two more stages, commonly termed “all red”, may be included to ensure that the junction is cleared before a conflicting flow begins. The sequence is shown in

Table 2.2: Light Sequence with Safety “All Red” Stages

|  |  |  |
| --- | --- | --- |
| Stage | NS (SN) light | EW (WE) light |
| 1 | Red | Red |
| 2 | Red | Green |
| 3 | Red | Red |
| 4 | Green | Red |

While the all red stages are technically not required in simulation systems they are included in realistic simulations to allow clearance of junctions and because they can have a significant impact on flow dynamics. They also provide a useful “landing point” for the traffic controller under development. Finally, for practical purposes and to assist human response times, yellow (or amber) stages are also included, which makes a total of 8 stages for a simple junction as shown by . In the rest of this report yellow and amber are used interchangeably. While it is common to refer to the caution light as amber in the UK, it may be more readily translated into other languages if yellow is used.

Table 2.3: Full Sequence for Simple Junction

|  |  |  |
| --- | --- | --- |
| Phase | NS (SN) light | EW (WE) light |
| 1 | Red | Red |
| 2 | Red | Red / Yellow |
| 3 | Red | Green |
| 4 | Red | Yellow |
| 5 | Red | Red |
| 6 | Red / Yellow | Red |
| 7 | Green | Red |
| 8 | Yellow | Red |

For calculation purposes, the yellow stage times can be included as part of another stage, for example the yellow of stage 4 can be included as part of the green of phase 3 or, more correctly, as a part of stage 5 / 6.

For clarity:

* The “effective green time” is the sum of phase 3 and phase 4 minus a period of time at the start of phase 3 to allow for human response time and a period at the end of phase 4 to ensure compliance with the red signal. The effective green time is dependent on human reaction times and driving styles and therefore is not clearly defined or bounded.
* The intergreen period is the time between consecutive green lights; in Table 2.3, this would be the time for phases 4, 5 and 6.
* The cycle time is the total time taken for a complete sequence to repeat.
* Split time is the percentage of the cycle time that the light shows green.
* Some other definitions are given in 0

Sequencing and timing become complicated for many reasons including when the junction is not physically and logically symmetric. The lack of symmetry can be the result of independent left and/or right turn signal, contains more (or less) than four approaches, includes a mix of one-way and two-way legs, has an interaction with another junction, etc.

### Detectors

Traffic detection can be based on a wide range of technologies, each of which has benefits, advantages and limitations.

Common detection methods can include magnetic, inductive loop, video (visible and infra-red), laser and radar [28]. Traffic detection includes class of vehicle, approach speed, vehicle distance, traffic volume and traffic flow.

Multiple detectors may be required to provide detailed flow information. For example, unmodulated continuous wave radar uses Doppler Shift to determine direct line-of-sight vehicle speed, it cannot measure distance to target, perpendicular speed or presence of slow and stationary objects. It is not possible for this type of radar to detect stationary queued vehicles and as a result it cannot measure queue length. Another form of detection would be required to determine the presence of a queue.

### Control

#### Fixed Timing

These are the simplest controls to consider. Traffic lights which are only capable of timed control are obsolete having been replaced by VA systems. The light sequence and timing are both fixed and as a result there is no need for traffic detection. To be effective, the controller needs to be tuned to the actual roadway in terms of cycle time and split, giving priority to the main route if this exists. As traffic trends change over time, the tuning will become dated and the junction will operate below the theoretical optimal rate.

A major failing of timed systems is related to the inability to react to dynamic traffic flow volumes. The timing necessary to provide adequate responsiveness while maximising throughput balanced across at least two competing flow directions has to be calculated in advance. The calculation is based on physical vehicle counts collected by observation over a long period of time and so fixed timing does not respond to unanticipated “events”.

Fixed timing is very inefficient when the traffic flow rates are low. When flow rates are low, the probability of a vehicle arriving is low but the random nature of traffic means the probability of the vehicle arriving during a red phase is over 50% (assuming an equal split). There is an entirely even probability of the vehicle arriving at any time during the red phase. Overall, a vehicle is more likely to experience a delay than it is to avoid a delay. Flow losses due to signal changes can be insignificant because of the low flow rates but an individual vehicle can be delayed by longer than a full red cycle time if it arrives at the signal as the signal transitions to yellow; the vehicle will then wait for the combined yellow, red and red/yellow periods. The low flow rate also implies that the probability of a vehicle arriving at a conflicting direction is also very low therefore the initial vehicle may wait without reason. The minimal number of vehicles is unlikely to require the entire green time in any direction.

By contrast, fixed timing can be the most efficient method when junctions are over saturated as the number of signal changes, and hence the flow losses due to signal changes, are minimised. The implication of a congested junction is that traffic is waiting in all directions and because of this, each green stage should be fully utilised.

Simple fixed time control is not in common use, possibly due to the cost/benefit advantages of installing VA systems. Other more advanced systems can be forced to operate in fixed timing mode where this is found to be advantageous. VA systems revert to an equivalent to fixed timing during congestion.

Isolated, fixed timing junctions with 50% split which are busy but not congested are frequently used as the benchmark to measure the improvement of proposed flow algorithms.

#### Vehicle Actuated

Vehicle Actuated (VA) systems resolve some of the problems of fixed timing by detecting traffic approaching the signal and responding accordingly. Vehicles have the same probability of arriving at a red indication as for timed systems but “calling for green” reduces the time delay experienced when there is no conflicting flow. Original VA systems detected traffic at the stop line, changing to green when required. Traffic lights change in the normal sequence and with a minimum green period of typically 7 seconds. During the green phase, a vehicle detected on the approach causes an extension of the green period. The extension may continue up to a maximum timer even if the flow rate is relatively low, which deprives the conflicting traffic of flow priority and may seriously impact junction efficiency. The detectors are set to “fail active” [29], meaning that a lost signal is the same as a vehicle detected signal; failures may cause unnecessary green extension in a single direction and a correspondingly unnecessary red in the conflicting direction. For flow efficiency and to prevent instability, VA systems prioritise existing flows over new requests.

Microprocessor Optimised VA (MOVA) improved the efficiency response and fault tolerance of the signals by using two detectors as shown in Figure 2.2.

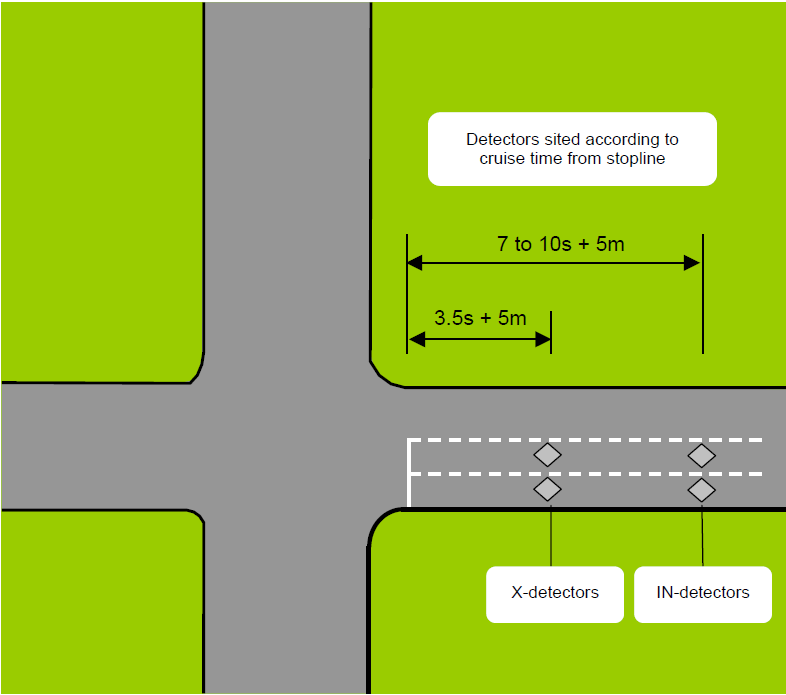


Figure .2: MOVA Detector Placing. From MCH1642 iss. C

An approaching vehicle will trigger a signal change request before actually stopping. A minimum green timer prevents rapid transitions while a maximum green timer prevents any direction from becoming permanent green due to either flow or detector failure. Detection is commonly a road inductive loop however it is not limited to this type, methods include vehicle presence, vehicle count, vehicle speed, vehicle type.

MOVA operates on isolated junctions though there is communication scheme known as “linked MOVA” available for where junctions are too closely spaced to be considered isolated. Despite the intention to operate MOVA only at isolated junctions, section 2.5.1 of the MOVA Installation Guide [30] makes reference to cycle time control to avoid exit blocking. The MOVA traffic Control Manual [31] states that there are circumstances where MOVA is ineffective; the description clearly indicates junctions which are not isolated. Full explanations of all operation and features are available in publicly availably documentation such as ”MCH1542, Installation Guide for MOVA” [30], “MOVA Traffic Control Manual” [31] and “Traffic Advisory Leaflet 1/06” [32]. The following is an abbreviated theory of operation taken from those publications.

MOVA has a predetermined sequence of phases. There are two detectors for an approach. The “X-detector” is sited at a distance of (3.5 seconds at normal cruise speed + 5 metres) before the stop line and the “IN-detector” is (nominally 7 -10 seconds plus 5 metres) before the stop line. Each green has a minimum display time, usually 7 seconds. During a green phase, traffic flow is detected on the approach. If free flowing, continuous traffic is detected then the green phase is extended. If free flowing traffic is not detected, the green expires after a period of time which allows the last detected vehicle to travel from the X detector to the junction at “cruise speed”. During the green phase, vehicles are counted on the approaches to opposing directions (red phases). MOVA estimates the difference between the overall benefit of extending the existing green phase over the delays incurred to the opposing traffic. If there is a nett benefit, the green is extended; if there is a nett loss the green is terminated. This mode of operation is intended to minimise stops and delays. If the queue is not cleared by the end of the green period, the approach is determined to be over-saturated and the junction moves to a maximising capacity mode of operation. Saturation is measured after the start of a red phase. A lane is deemed to be over-saturated if either the queue rate of growth exceeds a predetermined level or the IN-detector has continuously detected a vehicle for 6 seconds. MOVA allocates a longer green time to oversaturated links to minimise the lost time due to phase changes. There is an efficiency calculation based on a double differential of the change of flow; the rate of change of the rate of change of flow is used to determine if the queue is clearing. The green is terminated dependant on the efficiency calculation or a maximum green time. The detector information during oversaturation is not suitable to calculate stops and delays. Where there are under- and over-saturated approaches to a junction, the under-saturated approach methodology is to clear the queue and not apply any optimisation.

Based on the theory of operation, deficiencies of MOVA are that it only detects approach traffic. During a green phase it is possible to determine that there is congestion after the junction but only when the queue length is beyond the X-detector. There is no method to determine downstream congestion at a red signal and hence determine the benefit or otherwise of changing to a next phase. If the change is to a phase which cannot offer feasible flow, there is a traffic imposed artificial de-facto all red (total loss) phase which lasts for the minimum green time (7 seconds) plus the inter-green (all-red plus yellow plus red/yellow) time, equating to a minimum of 12 seconds total loss. The length of queue cannot be determined beyond the distance to the IN-detector, saturation (congestion) is estimated based on arrival model during the red phase. Only green signals can be demanded, there is no method to demand a red signal. MOVA falls back to an effective timed control when any approach is deemed over-saturated; the oversaturated approach moves to a maximum duration green timer whereas the under-saturated approaches move towards a minimum duration green timer.

There are several aspects of VA systems that limit their performance e.g. if traffic flow is continual from every direction each signal will run to respective max green times which is the same effect as fixed timing; the max green timers need to be tuned for the specific junction. In mitigation, several base timing tables can be made available and selected according to time of day, day of week, regular special events, etc.

While actively managing flows, VA controls adjust the length of the green time period during the current cycle; signals for conflicting flows adjust the green time in the conflicting directions thereby changing the red time in the direction under consideration. An important effect of this is that the cycle time is not a fixed period and will continually vary in accordance with traffic demand. Without absolute and precise traffic flow information, it is not possible to predict the state of a VA signal at any point in the future. The ability to use formal mathematical methods to investigate traffic controls may be limited as a result.

#### Coordinated Signals

Traffic flow through two or more isolated signals located within a short distance is likely to interact with both signals and experience up to n delays where n is the number of signals. In order to gain advantage from the interaction, coordinated signals use an offset time to determine when to transition. The offset time should be set to the time required for a vehicle to travel from the first signal to the second signal, there is then a further offset to the next signal and so on. The objective is to form a “green wave” along the primary flow direction with a platoon of vehicles passing each signal without stopping. Flow is triggered by the first signal in the set which may be timed or AV.

The operation is illustrated in Figure 2.3 and Table 2.4, which show a hypothetical set of six signals with different offsets due to physical distance differences. The cycle time is set to 90 seconds and all values are seconds except the signal number. Assuming that the vehicle arrives at signal 0 as the signal changes to green, each consecutive signal should also be green as the vehicle approaches at normal speed. In this idealised solution, vehicles experience only the delay at signal 0, which is zero in the example. Every vehicle in the platoon which passes through signal 0 in any cycle should be able to pass through each signal without delay.

Table .4: Vehicle Flow Through Six Coordinated Signals

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| signal (n) | 0 | 1 | 2 | 3 | 4 | 5 |
| offset to n+1 | 10 | 12 | 8 | 15 | 5 |  |
| vehicle arrives at | 0 | 10 | 22 | 30 | 45 | 50 |
| signal at arrival |  |  |  |  |  |  |
| next green | 0 | 10 | 22 | 30 | 45 | 50 |
| green (0) | 0 | 10 | 22 | 30 | 45 | 50 |
| red (0) | 45 | 55 | 67 | 75 | 90 | 95 |
| green (1) | 90 | 100 | 112 | 120 | 135 | 140 |
| red (1) | 135 | 145 | 157 | 165 | 180 | 185 |
| green (2) | 180 | 190 | 202 | 210 | 225 | 230 |

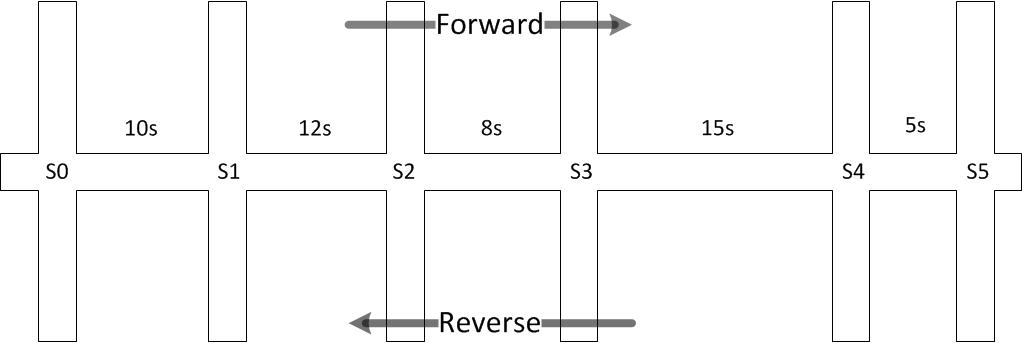


Figure .3: Signal Coordination across Six Junctions

The limitations of coordinated signals become apparent when vehicles attempt to use the route in the opposite direction. The situation is presented in Table 2.5. In this case a vehicle arrives at signal 5 at the exact same time as the vehicle arrived at signal 0 in the previous example. The vehicle will arrive during the green of the cycle before the example begins as indicated by green(-1). Travelling at normal speed, the vehicle arrives while signal 4 is showing red and is forced to wait 40 seconds; there is a similar problem at signals 2 and 0. In this example travel in the forward direction takes 50 seconds but 180 seconds in the reverse direction. Not readily apparent is the fact that any platoons will be broken because not all vehicles will be able to pass in any cycle e.g. the fourth vehicle arriving at signal 5 will be delayed by 6 seconds and will arrive during a red signal. In the forward direction 22 vehicles would all pass as a platoon.

Table .5: Reverse Flow at Coordinated Signals

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| signal (n) | 5 | 4 | 3 | 2 | 1 | 0 |
| offset to n-1 | 5 | 15 | 8 | 12 | 10 |  |
| vehicle arrives at | 0 | 5 | 60 | 68 | 134 | 144 |
| signal at arrival |  |  |  |  |  |  |
| next green | -40 | 45 | 30 | 122 | 100 | 180 |
| green (-1) | -40 | -45 | -60 | -68 | -80 | -90 |
| red (-1) | 5 | 0 | -15 | -23 | -35 | -45 |
| green (0) | 50 | 45 | 30 | 22 | 10 | 0 |
| red (0) | 95 | 90 | 75 | 67 | 55 | 45 |
| green (1) | 140 | 135 | 120 | 112 | 100 | 90 |
| red (1) | 185 | 180 | 165 | 157 | 145 | 135 |
| green (2) | 230 | 225 | 210 | 202 | 190 | 180 |

A related problem is that traffic is generally tidal, the major flow will be in a direction in the morning peak and the opposite direction in the afternoon peak; there will also be a period where there is no major flow. Determining the priority route is challenging and does not respond to unexpected peaks caused by other factors such as diversions from nearby routes.

There is an assumption that the offsets will be adequate to clear the junctions by the end of the cycle however this assumption could be flawed due to inconsistency in driver reactions or traffic joining the main flow from one of the conflicting directions.

Priority is absolute in the flow, there is no consideration of the conflicting or opposing traffic queues. Queues forming in any other direction will be serviced when the priority flow allows.

Co-ordinated signals need to be tuned as previously discussed.

#### Adaptive Urban Traffic Controls (UTC) general

UTC systems are intended to resolve many of the limitations of earlier systems by combining traffic signals into groups to improve traffic flow in an area or along a corridor. There are several variations on the theme including UTOPIA (Urban traffic optimisation by integrated automation) [27], SCATS (Sydney Coordinated Adaptive Traffic System) [33] and SCOOT (Split Cycle Offset Optimisation Technique). Two of the most popular are systems are SCOOT and SCATS. A common feature to these systems is the ability to modify the split and cycle times independently i.e. the split percentage can be kept constant while the cycle length changes or the split percentage can change within a fixed cycle length. In order to achieve this, the timing is modified after a cycle completes and not during the current cycle as is the case with VA. The offset period between consecutive lights is also variable. To prevent wide fluctuation in signal timings, the proposed changes are confirmed over several cycles before being applied and timing changes are limited to a few percent in any consecutive cycle meaning that several cycles may be required to apply the entire change. There is therefore an element of prediction required to determine the response that will be necessary a few cycles after the detected event. Adaptive controls are typically continually self-tuning based on rolling historic data such as the flows in the previous hour, at this time on the previous day, this day last week etc. The dynamic response of UTC may be impacted by filtering and level of change limitation. The application of previously determined models and the reduced dynamic response can minimise the ability of UTC to resolve unexpected and rapidly changing traffic flows.

#### SCOOT

In simple terms, SCOOT counts upstream vehicles that are approaching the control point. By recording the speed and number of vehicles, predictions are made regarding queue lengths and when phase changes should be made. This system relies on traffic flowing at the detection point, congestion is recognised when traffic is stationary at the detector [34] [35].

SCOOT is a complex, computer controlled Urban Traffic Control (UTC) system which is designed to operate over multiple junctions (nodes) in multiple regions. SCOOT applies the results of detected traffic to modelling system with the intention to estimate and predict queue lengths and dispersals. Estimation allows the system to respond to the current situation whereas prediction allows the system to prepare for future trends [36].

The basic operation of SCOOT is to detect traffic, model outcomes, compare to actual flow and apply the most effective control strategy. The following information is gathered and collated from a variety of sources and presentations including [35] [37] [34] [38]

SCOOT uses detection which is placed a distance upstream from the stop line, preferably immediately after the previous traffic signal and beyond the distance that will be occupied by the longest queue that is expected to form. The detected data is collected and converted to link profile units (lpu) which combines aspects of flow and occupancy. The data over a period is compiled into a cyclic flow profile (CFP) and patterns are extracted to determine, for example, platoon formation. Metrics used to measure control effectiveness include Degree of Saturation (DegS) and Performance Index (PI). Traffic controls within a region operate with the same cycle period to maximise the flow. Region boundaries are predetermined and occur where the distance between control signals is considered too great to allow effective co-ordination.

SCOOT has three optimisation systems one for split, one for cycle and one for offset, each of which estimate the overall effect of small incremental changes in each attribute.

The split optimiser updates before the phase change of every controlled junction. The objective is to determine if traffic flow would be most efficient with the current split time or with more or less green time. Efficiency in this case is determined to be minimising the maximum degree of saturation. The maximum amount of change is predefined and normally in the range 1 – 4 seconds.

The offset optimiser is run each cycle for every approach. Using modelled results and the CFP, the optimiser attempts to improve the platoon (or green wave) c34apability of consecutive junctions by adjusting the time differential of changes between upstream and downstream junctions. The change is varied in 4 second increments.

The cycle optimiser is least frequently run, after every 2.5 or 5 minutes. The optimiser identifies a single junction in each region as being critical and attempts to adjust the cycle time to maintain a 90% degree of saturation on each approach. Changes can either increase or decrease in 4, 8 or 16 second increments. Because the cycle time is consistent for all junctions within a region, any cycle time changes affect all junctions in that region. The cycle length increases in direct proportion to the traffic demand up to a maximum value.

Congestion management is achieved by gating (preventing further traffic from entering a region). This has been described as being “*…used to “relocate” traffic queues to links where they do not affect critical approaches to junctions…*” [39]

Other functions include prioritisation of specific vehicles, e.g. normal public transport, late (to schedule) public transport. There is also facility for “emergency green waves”.

##### Limitations of SCOOT

There are many claims made regarding the benefits of SCOOT however it is clear that the system is not perfect and does have limitations.

Reliance on solution models derived from earlier traffic flows is only viable while the current flows conform to previously identified situations. By definition, an exceptional circumstance will not have a model response and therefore a correct response cannot be applied. The rolling update model system is intended to smooth out random events, removing wide variation; the response to the exceptional circumstance will be diluted over a period of time until it is no longer available even if the same event re-occurs.

SCOOT does not normally automatically skip phases or change sequence. Phase skipping is available for prioritising busses when required however prioritising a single low frequency class does not resolve congestion, it creates additional delays for the higher frequency classes i.e. passenger cars. The prioritisation system can only provide benefit while the priority traffic is the minority class. Priority systems may require additional lanes to be built or a restriction of the road space available to other vehicle classes.

The system does not actively resolve existing congestion. The application of gating reduces congestion within a specified region at the detriment of the surrounding approach regions. Gating does not prevent or resolve congestion, simply moves congestion to a “more politically suitable” location.

In common with all approach detection systems, the detection becomes ineffective when queueing extends back to detector. The inference is an infinite queue and this cannot be efficiently resolved. If an infinite queue is identified on more than one approach, a fair sequence cannot be identified and there will be a reliance on manually collected turn data.

The upstream location of the detectors means that traffic sinks and/or sources that exist between the detection and control point cannot be managed. This could be a particularly difficult problem to resolve when there are temporary sinks and sources e.g. roadside retailers. The problem is caused when a vehicle is counted entering the controlled route but then temporarily stops inside the route and therefore does not pass during the modelled cycle; this is an inefficient use of the upstream control sequence. When the vehicle continues the journey, it will reinsert within a cycle that it has not been counted for which may result in queueing at a downstream signal.

The location of the detectors also means that individual lane detection is pointless; vehicles can and will change lanes in the approach to the signals. The proportionate allocation of time at a traffic signal requires previously collected turn data which may become stale over time or incorrect during transient flow periods.

Cycle time optimisation for the entire region is based on the perceived best solution for a single “critical” junction in the region. This might not be optimal for all junctions in the region especially during the congestion build up and clearance phases. The optimisation attempts to prioritise a single green wave path with the aim of resolving a predicted flow volume and direction.

A very important observation regarding the detection and update sequence is that SCOOT cannot change the current cycle if congestion develops. Once the current response has been selected, several cycles must pass before a further response can be applied.

#### SCATS

In similarly simple terms, SCATS detects vehicles at the stop line. Measuring flow rates during the green phase and the waiting time at the stop line, it is possible to estimate the queue length upstream from the control point. The SCATS detectors can also provide individual lane detection allowing “phase skipping” when no traffic is detected [40].

SCATS calculates a best local solution based on current detection volumes. When adjacent signals agree on the split and cycle periods, they undergo a marriage and synchronise the signals. At some later stage, if adjacent signals disagree over timing the signals undergo a divorce. The marriages and divorces system allows a degree of flexibility to determine and respond to the congestion region boundary.

##### Limitations of SCATS

SCATS uses actual traffic measurements to determine a suitable response profile rather than using a modelling approach. The detectors are closer to the lane ends and therefore per lane detection is used. However SCATS determines the next suitable change over a period of cycles and is limited by the change in split and cycle that is applied on each consecutive cycle. In common with SCOOT, SCATS is therefore making a prediction regarding the solution that needs to be applied several minutes after the actual detection. The solution will have transient peaks and troughs smoothed out by the integrating action of multi-cycle averaging making the system less responsive to highly dynamic changes

#### Intelligent Transport Systems

Future detection and reporting methods will include the use of in-vehicle sensors, V2V / V2I / V2X communications and VANET (Vehicular Ad Hoc Networks) [41] [42] [43]. The exact function of the detection and reporting is not fully defined but it is assumed that the detection will in effect be a form of approach detection and there will be some form of third party arbitration and signalling.

#### General Limitations of Existing Methods

Existing systems have a reliance on some form of flow and/or timing prediction. Actuated solutions use detection in advance of the control signal, transition decisions are based on a single input data source of approach data. Systems make an assumption that traffic flow is always possible and have an inability to confirm the suitability of the next signal activation. These factors become significantly problematic as the traffic flow rate increases.

Demand is made only for green time, opposing flows transition to red as a pre-emptive consequence.

Upstream detectors become saturated and are then ineffective as a data source. Theoretically a traffic queue could actually be an infinite length and therefore no upstream detector can detect the total length of a queue. Moreover, there is also a critical detection distance where a queue length cannot be satisfied within a traffic light cycle even when the signal timing is at the extreme service condition, e.g. minimum cycle time to satisfy the condition of maximum green time in combination with legal minimum red time.

Traditional “in surface” loop detectors can only detect traffic at the single point of insertion. There is a potential for vehicles to stop either side of loop detectors and not be registered. Likewise, queues which fall short of the detector by less than a metre may not be registered, whereas queues which extend to the detector may be registered as infinite length. A change in queue by less than the length of a single vehicle can therefore have far reaching consequences for the signal timing. The solution to this problem is not related to alternative detectors as each type of detection system has some finite range and object discrimination capability limitation.

Detection of stationary vehicles can be especially problematic. If a stationary vehicle is not directly over an in-surface detector it cannot be detected. Similarly, stationary vehicles might be undetectable to Doppler Shift radar systems (frequency difference between transmitted and received signals) which rely on movement to produce the Doppler Shift effect.

Of prime interest in this work is the limitation of detecting only traffic approaching a junction as opposed to traffic departing from the junction. Traffic detected at the approach to a junction is necessary to determine optimal sequencing and fair allocation however this presupposes that the vehicles are able to comply with the signal indications. Vehicles will only be able to comply if there is available capacity in front of their current position; at a junction this means that the intended exit must be clear. For current traffic control systems a clear exit is assumed but cannot be proven. Under normal circumstances, the assumption will be correct and traffic will move as directed but in congestion there is a significant probability that exits may not be clear and therefore traffic may not be able to comply with the signalling intent. The blocked exit will increase the length of the approach queue, potentially propagating congestion back upstream to affect other signals. Without an exit detector available exit capacity availability can only be determined by flow detection. Flow detection is only possible if the light is green and the exit is clear. The availability of capacity cannot be determined during a red signal and a needless transition could occur potentially stopping a conflicting flow and increasing the probability of congestion in the conflicting direction. It is theoretically possible to detect an exit becoming blocked during an indicated green using approach detection but the location of the detectors adds a significant delay to detection. For example, a MOVA detector is usually positioned at a distance equivalent to approximately 5 seconds of travel before the stop line at normal approach speed and so there is a guaranteed minimum delay of 5 seconds before a block is detectable. The distance between a SCOOT detector and the signal could be measured in minutes.

### Predictive Methods

The effectiveness of a traffic control system is based upon satisfying the greatest overall number of vehicles and presenting the minimum overall delay. To this end, a prediction needs to be made regarding how the junction will be used. The prediction may be one of two forms:

1) long term planning: predicting the viability of a change to existing control

2) short term response: predicting the traffic arrival at a junction in the next few minutes / hours

The relevance of both aspects are briefly discussed in this section

1) Planning: Prior to installing a traffic control system, a measurement needs to be made of the actual traffic flow at the location to be controlled. Traffic volume and localised source-destination information is collected to determine the most suitable time distribution across the junction in question. This timing and turn information forms the basis of the initial state signal timing. For timed signals, it is the actual state transition timing, for VA signals it is used to set the maximum times and for adaptive signals it is used to set initial timing and to determine sub-area boundaries. A major factor that needs to be considered is the effect that the new timing will have on traffic flows; any direction at any time could become more popular or less popular as a result of the timing change. The change of popularity changes the number and rate of vehicle movements which in turn affects the desired signal timing.

It is an assumption that the cost of implementing or modifying traffic control measures will only be borne when it is proven that some form of flow problem exists or is likely to exist after a junction is created or substantially modified.

For a new or substantially modified junction no flow data can be collected and the timing can only be estimated. However, for an existing junction, a flow problem might imply that there are excess delays caused in one or more directions by inefficient control or inefficient sequencing. The existing control method employed may or may not be active control, e.g. the junction may use traffic lights, stop or give-way markings or be some form of traffic island.

There is potential that that any data collected may misrepresent desired or ultimate flows, e.g. a percentage of drivers may use an alternative route to avoid the congested junction with the aim of reduced travel time even though this is at the expense of travelling additional distance. The result is that the collected flow data does not include all of the potential “normal” traffic as some traffic is taking a suboptimal route. When the junction change is completed, traffic may attempt to use the more direct route, increasing flow density and consequently causing inefficient operation. The inefficiency may be relatively minor, causing suboptimal operation or significant and causing greater delays than previously experienced. This effect is, in many ways, similar to the theory of induced travel. The theory has been quoted many times although a specific attribution is difficult to identify. Noland [44] cites several attributions including dates to 1938. In simple terms, the theory states that if a destination is popular then demand will exceed access capacity and there will be a queue. Logically speaking, if the destination popularity remains constant and access is improved, the queue should reduce. However, with the same popularity and better access the overall attractiveness of the destination increases because the deterrent effect of queueing has reduced. The result is that the destination becomes better attended, the increased patronage utilising the increased capacity and the queues return to the same level as before the access was improved. In transport terms, a road section is congested because it provides a useful mass route to a destination; increasing the road capacity enables more vehicles to benefit from the route. This concept can also be explained as Braess Paradox [45] [46]. The “anti-paradox” to this is a concept where the increased patronage results in a perception of reduced service, overcrowding, degradation of facilities or lack of exclusivity.

2) Response: As explained in sections 2.3.4.5 and 2.3.4.6, UTC systems detect flows and apply suitable rules several cycles later. The rules applied relay on prediction of immediate solutions based on historic events. However, route use is neither consistent nor constant; the choice of route and time of presence for any given user is variable with wide ranging dependencies including work load, personal appointments, social events, previous route section delays, vehicle condition, weather conditions, holiday entitlement and so on.

The effect of accuracy of prediction of flow efficiency becomes clear when considering that queueing begins with a single vehicle. At traffic light controlled junctions, queueing begins when the single vehicle does not pass through the junction within the current cycle. Given that a junction will have some vehicle capacity, Vc, averaged over the cycle time; it will also have a number of vehicle arriving over the cycle time Va.

Equ( ‑)

Equ( ‑)

Equ( ‑)

Equ( ‑)

At the end of the cycle time, t, the queue length at the junction will be unchanged for all cases if Equ( 2‑1) is true. If Equ( 2‑2) is true, the final queue will be longer than the initial queue in all cases. The queue will reduce if Equ( 2‑3) is true and the initial queue was > 0, otherwise the queue will remain at zero length and the junction may be operating inefficiently by allocating unnecessary green in the direction under consideration but denying potential green time from being allocated to the conflicting direction. An important aspect at this point is that the values of Vc and Va are averages and as a result, although they represent vehicles, they do not need to be integers. However, when applied to an actual junction flow, the fractional parts become important; fractions of a vehicle cannot pass the junction, only complete vehicles. The queue will increase if Va is only fractionally larger than Vc. It also becomes clear that the queue will not change if either Equ( 2‑1) or Equ( 2‑4) is true; in this condition the junction is at its most efficient when the direction is taken in isolation. To be efficient and avoid queues, the prediction needs to be accurate within one vehicle of the actual flow. Under normal circumstances this is a two second window of opportunity.

There is a transition phase from free flowing to congested traffic where traffic flow is not significantly better with timed or actuated lights. There is some critical value where a changeover takes place. The critical value for increasing flow and a change from actuated to timed control is higher than the critical value for decreasing traffic flow, there is some natural hysteresis loop.

The accuracy of the prediction model, or rather the correlation with physical measurement, has greater significance over the traffic flow and levels of congestion as the flow of traffic reaches the critical level. While the flow level is below the critical value, small errors in prediction are naturally resolved by the slow rates of called transition; queues are cleared in each stage. When the presented traffic flow is above the saturation level, residual queues will form regardless of small errors in the prediction. However, at the critical traffic level, small errors will give rise to the beginning of queue formation in one of the opposing directions which will rapidly propagate.

* 1. Traffic Flow Optimisation Techniques

In order to determine the outcome of any traffic situation, there are two main options. The first is to physically test the scenario using the actual road and vehicles. This would be challenging for all but the least complex scenarios, requiring precise vehicle vector and control system information. Even when the detail of the main systems are known, other factors which are more difficult to control, e.g. road temperature or wind direction, may have an impact. The second alternative is to predict the outcome based on mathematical models of the constituent systems. This is typically lower cost, ensures that input factors are controlled and gives a high degree of repeatability. However the results are based on the accuracy of the models employed. In terms of vehicle modelling, it is possible to specify mechanical limits such as rates of acceleration or minimum distance to the preceding vehicle but not possible to ensure that every (human) driver will adhere to these limits at all times. In either case, there are factors which are accurately controllable and factors that are not. The impact of this is explained in Chaos Theory.

### Chaos Theory

Chaos Theory was defined by Lorenz in 1963. The seminal paper discusses the effects of dissipative hydrodynamic flow and the feasibility of long range weather forecasting [47]. Lorenz is variously listed as originating the quotation “The approximate present does not approximately determine the future.” This simple statement has an impact in many fields and especially in predictive traffic management.

Chaos theory argues that the future can only be accurately predicted if and only if all variables and constants are accurately defined, measured and applied at the beginning of a sequence of events, these are commonly known as the set of initial conditions. The effect of minor variations on continuously linear events are cumulative, becoming amplified over successive time periods implying that the reliability of a prediction is inversely related to the length of time that the prediction is made for. Lorenz original paper describes non-periodic functions, those that do not contain absolute periodic or cyclic reset (or synchronisation) features. A periodic reset creates a starting datum at regular intervals; the prediction value will also reset at this point. For a cyclic system with a periodic time of x, a prediction made after a period of 5, x+5, 2x+5 should always return the same result; it is assumed, but not imperative, that x and 5 are measured in the same scale and that x > 5. While macro traffic flow can in some ways be considered as periodic, micro flow is certainly not periodic

Two items of significance are the included variables and the accuracy of representation. Determination of all variables may not possible for a system which is unbounded, there is always potential for an undefined or unexpected impulse input in such cases. The unexpected input is generated by some uncontrolled source such as a pedestrian, small animal or climatically induced movement (wind assisted floating debris)

Accuracy of representation infers the ability to accurately determine a starting position but also necessitates accuracy of measurement and representation. The significance of Chaos Theory can be recognised by considering a measurement of e.g. 0.5 units of some measurement scale. This implies a measurement of 0.5 followed by an infinite number of zeros, each of the zeros being significant. If there is not an infinite number of zeros, the number following the last zero could be any number other than zero, and therefore the starting position is only an approximation, not an accurate representation. If the number is stored by some digital means, there will be a limit to the number of bits that can be held and this can allow a rounding accuracy. Furthermore, it is well known that not all numbers can be accurately represented in binary based systems, many fractions may only be approximate. For example, one third can never be accurately expressed in any other way than a fraction. There is a similar limitation for very large numbers that extend beyond the integer range of the digital storage. Finally, any angular representation which requires the use of the constant pi will only be an approximation as pi is understood to be an infinite string of digits that never repeat, there is always a number following the last number considered and again this is an approximation.

The significance of this relates to an inability to accurately specify the initial conditions for any individual vehicle. It is not possible to specify with absolute accuracy the time that the vehicle begins its journey or the acceleration rate from rest nor is it possible to confirm that all initial conditions are being considered (e.g. friction coefficient of road and tyre, wind direction and speed, visibility due to climatic precipitation and the indeterminate mind state of the driver). Moreover for any multiple vehicle system with control system integration, the initial conditions are all relative and errors are accumulative. While it is common to consider urban traffic flow to be a single macro system, it is a fact that it is actually a construction of many micro systems, all of which exhibit their own but interacting Chaotic behaviours.

A review paper by Adewumi, Kagamba and Alochukwu [48] discusses Chaotic Systems in relation to urban vehicular traffic, stating the sensitivity to initial conditions, the need for determinism (single predictable next state based on the existing state and raising the point that systems must be finitely bounded) and instability (predictions become less accurate over longer periods of time). The paper cites examples of liquid cooling to an “exquisite shape” however it fails to acknowledge that this is a change of state from liquid to solid. The authors also state an uncited example of Chaos Theory predicting a stock market crash every two decades beginning in 1987 based on Mandelbrot Fractal Hypothesis, a prediction which remains a single point graph to date. For interest, the two decade prediction is refuted by Johansen and Sornette [49] stating events of 1914-18, 1929 and 1987 as being outliers in the Dow-Jones index.

The initial conclusion of this is that unless Chaos Theory can be proven to be false regarding sensitivity to initial conditions, it is not possible to accurately predict the flow and travel time of any individual vehicle during a non-trivial journey. As a consequence of this, it is also not possible to accurately predict the macro flow of urban traffic.

The second conclusion is identified in a publication by Frazier and Kockleman [50] who conclude that the system may not be deterministic due to the “random agents” including humans and weather. If the system is not deterministic then Chaos Theory cannot apply. They also state that if rules and laws are applied to the human agents then Chaos Theory may apply.

### Queueing Theory

Queueing theory considers queue lengths and waiting times for items in a queue. The origin of the problem explanation is generally attributed to A Erlang [51] in reference to queueing of telephone calls. The work was based on Markov Chains introduced by AA Markov in 1906 but has been developed to include data applications such as transmission line capacity, rate of disk drive writes and so on. When used in this field, the theory considers the concept that data packets arrive at some rate λ, are held in an input buffer while waiting to be serviced by a server at rate μ before passing to the discharge or output medium. The flow rate is determined by λ when if the input is not saturated or by μ (and the discharge medium) when the input is saturated.

Queueing theory can be illustrated using the following diagram

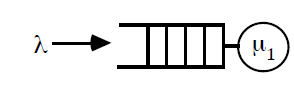


Figure 2.4: Components of Queueing Theory

The queueing is fully described using 6 component parts (T/X/C/K/P/Z) in Conventional (Kendall) Notation where:

T – Distribution of arrival times

X – Distribution of service times

C – Number of servers

K – Queue capacity

P – Size of population

Z – Service discipline (order of servicing)

For most purposes the queue capacity, K, and the population size, P, are considered to be infinite and the service discipline, Z, is defined as FIFO. These suggestions reduce the notation to 3 components T/X/C. The distributions (T and X) are usually regular patterns such as Markov (M), general (G), deterministic (D) or Erlang (Ek) although many other pattern descriptions are possible. C, the number of server elements is almost certainly an integer. The elements may also be identified as A/S/C to describe Arrivals, Servicing and Computing elements.

Traffic light systems at isolated junctions are frequently described as M/M/1 or Markov arrivals (no dependence on previous arrivals, Poisson distribution) / Markov servicing (service time is independent of any previous service time but Poisson variability based on vehicle type, size and acceleration capability) / single controlling element through the junction. Other important factors include having infinite input and output buffers (as in the case of [52]): the controller will hold any vehicle presented in the input queue until serviced, all serviced vehicles will leave the system without restriction. However, non-isolated junctions may be described as a single controller with finite input and/or output buffers: the input content of the input buffer is determined by the action of an upstream controller; the ability of a vehicle to leave the system is determined by the local controller and by any downstream controller. A single direction through a non-isolated junction can be represented by Figure 2.5

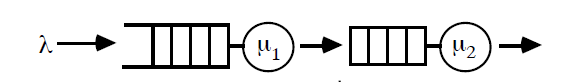


Figure 2.5: Queueing at a Non-isolated Junction

In the non-isolated case, the queue buffer for μ2 has finite capacity; the service rate of μ1 is dependent on the system input buffer, λ, the action of μ1, the input buffer to μ2 and the action of μ2. Similarly the service rate of μ2 is dependent on the system input buffer, λ, the action of μ1, the input buffer to μ2 and the action of μ2.

Optimising the throughput of a single direction through an isolated junction can be achieved by computing a suitable value of μ against λ. The calculation complexity is increased when conflicting flows with differing λ values are considered. In this case the size of the input buffer is also required. It is important to note that the balance of μ against λ is only possible while λ ≥ μ otherwise the queue will continually increase as suggested in Equ( 2‑1) - Equ( 2‑4). This is further explained using Littles Law which states

Equ( 2‑5)

Equ( 2‑6)

Where

Equ( 2‑7)

Also, given that

Equ( 2‑8)

It can be observed that the flow is balanced, i.e. the queue length will remain constant while

Equ( 2‑9)

When considering non-isolated junctions, as in the case of this work, it is possible to consider the input buffer of the next downstream junction, Jn+1 to be the output buffer of the junction under consideration, Jn. The combination of multiple conflicting flow and effective output buffers is shown in Figure 2.6.

Figure 2.6 when considered in conjunction with Figure 2.5 identifies a very important concept of this work; output buffers will be present throughout a route until some point where there is no downstream controller. At this point, which may be temporal or permanent, the system will be presented with an infinite sink capacity or the capability to remove any number of vehicles from the system. If the infinite sink points are identified at junction Jexit, the input buffer of junction Jexit (which is also the output buffer of Jexit-1) can be cleared which has the effect of moving the infinite capacity sink *backwards* into the congested region and thereby clearing the route.

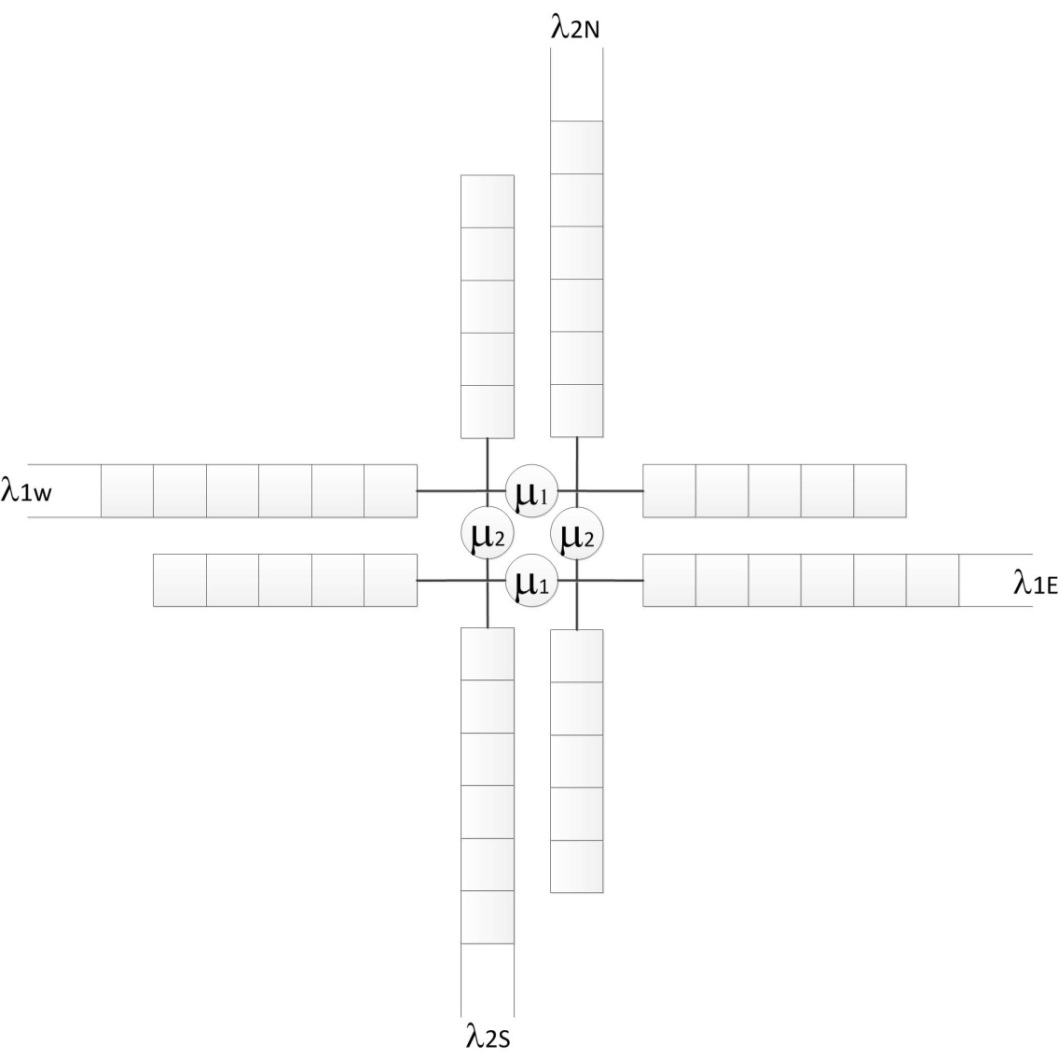


Figure 2.6: Queueing Theory Applied to Non-isolated Crossroad Junction

### Genetic Algorithms

Genetic algorithms are used as an attempt to find an optimal solution to a given complex problem. In a process similar to the theory of natural selection, (randomly) mutated children of a parent process are compared to the parent process to identify if they provide a “better” solution to an initial problem. If the solution is “better”, the children become parents and the new children are tested. The process continues until the “best” result has been found, this being apparent when repeated testing does not identify any “better” solution.

For illustration, consider two points in a maze, point P0 is the initial position and point S is the best solution. The distance between point P0 and point S can be determined however the location of point S is unknown and it cannot be directly observed from point P0. If a point Pn can be determined which is closer to point S, n is incremented and this becomes the new starting point Pn+1. The determination process is then repeated to identify another point Pn+1 which is again closer to S. Point Pn will be nearest to S when the determination process does not identify any point Pn+1 that is closer to S than the current point Pn. Where the determination process identifies more than one Pn+1 that is closer to S, the point that gives the greatest change is usually chosen; this is the greatest slope or highest rate of change per determination step.

From the maze illustration, there are several points of note:

* While the objective is to locate point S, this may not be possible in all cases; the final point Pn is then an optimal solution rather than a complete solution.
* There is also the possibility of a local solution being identified, where there is no point Pn+1 that reduces the distance to S from the current point Pn but where other points Pn exist which are closer to S. This is the equivalent to taking a dead-end turn in the maze. If the minimum achievable distance between Pn and S is known in advance then it is possible to restart the determination process from some earlier stage using different methods or measurements. This is only possible if the minimum distance (and hence the solution) is known in advance.
* The step size is important. Larger step sizes would generally give faster convergence however this may fail to identify better solutions. As an example, if the distance to target S is 5 units and the step size is 10 units there is no step that can be made that will improve on the current distance. If however the step size is reduced to 5 then an absolute solution can be identified. If the step size is reduced to 2 units, a better but not absolute solution can be identified. Smaller step sizes may or may not identify better solutions but will necessitate a higher number of iterations and hence greater time to complete.

As an alternative, the GA calculation can be applied to a process rather than a position. Consider a traffic control system. The origin and destination do not change but the length of queue can be changed by modifying the signal sequence and/or timing. In this case, queue is measured at some physical and temporal point and then the signalling adjusted. The difference in queue size during the next cycle is used to influence the next signal adjustment. Using this method it has been suggested that Genetic Algorithms could be used to find the timing sequence that will provide the maximum traffic flow at a junction. It should be clear however that this will rely on traffic arrival and clearance rates remaining constant otherwise there is no certainty that the queue has been influenced only by the signal timing change.

### Cellular Automata

Cellular Automata is originally attributed to Stanislaw Ulam and John von Neuman C.1940. While not originally derived or restricted to a single academic field, Cellular Automata is considered very effective in traffic simulation. It is used to describe the evolution the progress of an object (vehicle) over a series of finite time steps. The region of interest (in this case the roadway) is constructed from identically sized cells and the progression of the vehicle is dependent on the occupancy of the current and surrounding cells. For simple traffic systems, the cells are vehicle sized sections of a defined roadway that either do or do not contain a vehicle. At each time step, the vehicles may move in the direction of travel by a number of cells, the number is indicative of the speed that they are travelling. Vehicles are not permitted to move into a cell that is already occupied. In practice, a vehicle will occupy more than one cell, meaning that the cell size is smaller than a vehicle. There will be a required number of empty cells between occupied cells that represents the gap between vehicles, the number of empty cells is a function of the number of cells travelled at each step.

The difference in the number of cells moved in successive time steps represents the acceleration of a vehicle. Restricting or limiting the acceleration rate of change to practical values requires memory of the previous two locations to determine the next feasible location. A similar principle can also be applied to the rate of change of direction, i.e. any vehicle has a minimum turning circle, the number of cells required to complete a 90° change of direction is proportionate to the current speed and is always greater than 1. This is known as second-order cellular automata and is attributed to Edward Fredkin.

While the cell size is a fixed distance determined in advance, it is possible to consider the size of cell, number of empty cells and step time to be indicative of either “headway” or “occupancy”. Using this definition, the cells can represent a time difference and not a distance difference.

### Petrinets

Petrinets are a diagrammatical method of modelling the behaviour of state dependant systems and are especially suited to systems that have multiple concurrent state paths. They were originally created by C A Petri in 1939 with the purpose of describing and resolving the interactions in chemical processes. The Petrinet concept was published in 1962 [53] [54], developed over many years and adapted to other fields including computing and communication [55].

By knowing the current state and using inputs that were the outputs of the previous state, it is possible to determine the next state. For any system or process, a diagram can be drawn that indicates current states. The diagram is then repeated with updates that show new current states based on the previous state and changes. Using this method, areas of change can be easily identified and only relatively small changes made to an otherwise complex system. This can be useful for modelling and defining traffic light systems as it readily identifies states which either cannot be entered or cannot be exited and identifies spurious states that are generated when not required.

There are many alternative forms of Petrinets including “coloured” and “timed” versions. In Coloured Petrinets, highly complex systems that comprise of multiple different processes that do not directly interact can be investigated by considering the system state as a combined result of the individual unique processes, the “colour” of the input is the description of the discrimination. In Timed Petrinets, the transition to the next state happens after a fixed time period; this has the effect of synchronising state changes and/or forcing a transition to a new current state after the clock period even when there have been no other inputs.

Petrinets use symbols to diagrammatically identify aspects of a process; circles represent places/states, rectangles represent transitions, arrows connect the states and transitions (referred to as arcs) and a “dot” is used as a token to indicate the current state. A number next to an arc indicates a number of tokens required to follow the arc. A single traffic light head sequence is depicted as a Timed Petrinet in the following diagram as a means of clarification. The sequence passes the expected Red, Red/Yellow, Green, Yellow sequence, staying Red for 30 seconds, Red/Yellow for 2 seconds etc.

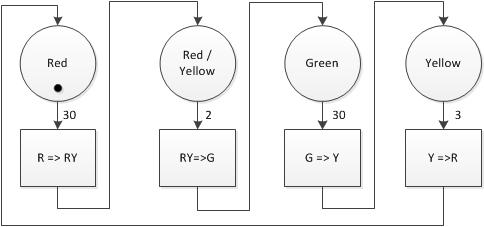


Figure 2.7: Example Timed Petrinet of Single Traffic Light Head

In a real system, this head would need to interact with the opposing traffic head for sequence coordination and in order to prevent a green from being shown in multiple directions. To ensure diagrammatic clarity this is not shown.

### Game Theory

Game theory is an extensive subject that has generated countless research papers and even Nobel Prizes for primary exponents such as John Nash, John Harsanyi and Reinhard Selten. The application of Game theory is very wide ranging and can include politics, economics, military deployment, disease control and so on.

Game theory is used to predict an outcome based on initial and transitional events. It is a wide ranging subject area that can also be described as interactive decision theory. Given the basic premise that any decisions made will be rational and provide the best individual outcome, it is possible to use game theory to predict the outcome of, or solution to, many different types of problem. It is possible, for example, to use game theory to predict whether a vehicle will overtake, exceed speed limits, stop for a yellow light or consider a distance to be safe. These attributes have a significant impact on congestion, queue forming and queue clearance in urban roadways and motorways.

There are many variations of Game Theory but they generally follow the same basic concept that an actor will undertake an action if the action is considered to provide a nett benefit. An underpinning principle is that the actors/players should be rational. A rational actor/player is one that will always choose the action that provides best result based on the information available at the time. There are theories that involve irrational actors/players to explore unanticipated outcomes but they are not being considered here.

There are several definitions of game types:

Sequential and simultaneous. In a sequential game, players make moves in order and are aware of all previous moves. The outcome of a sequential game is usually shown by tree structure. In a simultaneous game, players all make each move at the same time without knowing the other players moves. The inability of any player to take another action until all players have completed the previous action combined with a lack of knowledge about the other player actions effectively makes the moves simultaneous even when a relatively long period of time has elapsed. The outcome of a simultaneous game is commonly shown by a table. Chess would be an example of a sequential game, whereas “rock, paper, scissors” is a simultaneous game.

Zero sum and non-zero sum. In a zero sum game, the total value of all rewards is fixed. No player can gain any value greater than the sum of the other players losses. In other terms ∑wins + ∑(losses)=0. A obvious real example of a zero sum game would be poker though most games (in the traditional sense) are zero sum. Zero sum games are strictly competitive and because of this rarely represent real life where there are frequently conflicting and cooperative elements. In non-zero games a gain from one player does not necessarily result in a loss for the other and vice versa. A simple example of a non-zero sum game could be deciding who should have the last treat from the jar; the treat is not divisible, if all players choose to leave the treat to someone else the treat goes to waste. The fact any player did not receive the treat did not mean that another player would automatically receive it, therefore a loss by one player did not result in a win for another.

Co-operative and non-co-operative games. A co-operative game is where players commit to an agreement to secure an outcome. The agreement is most likely to be enforceable and contain penalty measures. The implication is that there will be communication in co-operative games. As zero sum games are strictly competitive and there is no mutually preferable outcome, they cannot be co-operative.

Symmetric games are ones where the outcome matrix is symmetric. In short, if player A uses strategy s, they will achieve outcome x; if player B used the same strategy, they too would achieve outcome x. In this way, the outcome depends on the strategy employed and is independent of the player that used it. Games where the outcome is dependent on player and strategy are asymmetric. An example asymmetric game would be a “cops and robbers” standoff, where only one player is likely to achieve a desirable outcome when employing a storming tactic.

Other important terms used in Game Theory:

Perfect informational. Where players are aware of all previous moves and outcomes (but not necessarily overall strategies) before committing to a move. Only sequential games can be considered to be perfect informational.

Nash Equilibrium. When, based on the position of the other players, no player can benefit from changing their position. This could be a stalemate situation or a position where all players are gaining the maximum benefit

Pareto efficient. When there is no outcome that makes a player better off without making another player worse off. Zero sum games are always intended to be pareto efficient. Nash Equilibrium does not imply absolute Pareto Efficiency.

There are many examples of attempts to empirically quantify transportation to allow the use of Game Theory as a flow design or prediction tool for example [56], [57]. However, in transportation and especially urban traffic there are many psychological influences which means, amongst other things, that the concepts of nett benefit are not fixed and this must be taken into account. An excellent example of this is provided by an ongoing discussion entitled “Viewpoints: How can the stigma of public transport as the ‘poor man's vehicle’ be overcome to enhance sustainability and climate change mitigation?” [58] in which varied economic, ecological and psychological arguments regarding private car ownership are debated. It is frequently concluded that private vehicle ownership is not cost effective however there are continuous counter arguments that the use of public transport is in some way socially demeaning. As a result of this, the real actors are making irrational decisions and therefore an over-reliance on Game Theory in transportation is questionable.

An indication of a changed concept of nett benefit can be observed where a relative minor road joins a relative major road. If the traffic flow in the major road is continuous even at a relatively modest speed, it is unlikely that drivers will give way to traffic attempting to enter from the minor road. If however the flow on the major road is very slow or periodically stopped, there is an increased likelihood that the driver on the major route will give way to vehicles attempting to join the route. According to Game Theory, and assuming the actors are rational, the drivers on the major route should consider the nett benefit of their own actions. In this case it would appear that drivers perceive a personal benefit of not losing journey time / distance or road position while the traffic is moving but perceive a nett personal moral benefit of politeness or due consideration of others at much lower speeds. Clearly there must be some critical speed where the drivers change between perceptions. In the context of this scenario, the personal moral benefit can be seen to be immediately depleted where another driver is felt to be taking advantage by entering when not invited.

An apparent opposite behaviour can be observed at a controlled junction where a driver may consider that a green signal is not allowing progression due to exit blocking. In this case a driver may move into the contested area knowing that they will not be able to clear the junction and thereby deliberately blocking the path of other traffic. In this case the driver/actor must perceive moving forward by a small distance to be a nett benefit above overall traffic flow.

### Simulation

The mathematical solutions presented in this section can be applied using manual methods to examine the predicted flow of traffic along a route. The normal drawback to this approach is the number of mundane repetitious mathematical iterations that are required to obtain a significant result. The speed and accuracy of the calculations can be improved using the automation features in common spreadsheet applications however these still have many limitations for example:

* Need for fixed, predefined inputs to the system
* Difficulty in identifying the minimum number of iterations required for effective convergence with corresponding rapid growth in spreadsheet length
* Exponential increase in spreadsheet dimensions (length, width and number of sheets) as the complexity of the system increases i.e. the number of interactions between vehicles and junctions
* Limited ability to visualise results
* Limited ability to confirm progress towards convergence

While other problems may be introduced, simulation can resolve many of these obstacles however modelling and the type of simulation must be understood from the outset. Simulation systems use mathematical models of physical entities to derive results. The models normally contain a list of parameters that relate to the functional or physical attributes of the entity. Where a highly accurate outcome is required from the simulation, the model will contain every known or derivable parameter to a very high accuracy and the simulation will run with a very small iteration value (normally this will be time). In direct contrast, an accurate simulation outcome can only be achieved where all the parameters are able to be accurately defined. These concepts can be considered alongside the Lorenz quotation from section 2.4.1 “The approximate present does not approximately determine the future.”

Where simulation systems are being used to determine absolute physical systems, high accuracy is normally required and simulation of active processing will be slower than real time. Such cases can include e.g. impact stress during landing of extra-terrestrial vehicles, deformation of crush zones during vehicle impact or changes in material properties due to heating in gas turbine blades. Accurate modelling and simulation of these systems can reduce failures, improve safety and minimise development time/cost.

Alternatively, simulation systems can be used to indicate a probable outcome based on known or anticipated parameters. In this type of simulation the iteration timestep is likely to be much higher and as a result the simulation will be much faster, frequently exceeding the realtime of the simulated process. The ability to use greater timestep is based on the inaccuracy of a next stage iteration derived from an approximated initial position affected by an assumed input. Traffic simulators are in this category due to the inability to accurately predict the behaviour of human drivers in any given physical circumstance. As a result traffic simulation is based on modelling uncertainty, managing probability and predicting outcomes.

Traffic simulation may use several of the previously defined concepts including Cellular Automata, Game Theory and Queueing Theory alongside vehicle operating parameter envelopes, road capacity calculations and probabilistic operator actions. The outcome accuracy can be enhanced by the inclusion of real world calibration data measured and collected in advance to provide a known outcome for known initial conditions; this allows the simulation parameters to be adapted to provide the same conclusion as the real world situation. Improvements can also be obtained by simulating and observing a live network in parallel in realtime, applying periodic corrections to the simulation while identifying causes for the deviation.

There are many conceptual differences between simulation for physical systems e.g. electronic circuit simulation and simulation system for psycho-physical systems e.g. traffic simulation. One of the first differences is that an input in physical simulation is a change that will cause a sequential change in the dynamic attributes of the physical system i.e. an increase of voltage would cause a proportionate increase in current flow under normal circumstances and the increase in current will cause a proportionate increase in temperature. Alternatively, in psycho-physical systems the input can be considered as an opportunity for a sequential change rather than a certainty of effect i.e. a traffic signal changing from red to green invites a waiting vehicle to move, the vehicle may move at any future timestep or may not move at all, if the vehicle does move it can accelerate at any rate from near zero m/s2 to the maximum acceleration rate of the fastest known road vehicle; it may even accelerate negatively.

When simulating physical systems it is normal to define precise, identical responses to a given input providing an output that can be relied upon in the real world, subject to the constraints of the model employed. Variability in component attributes or dimensions is introduced by schemes such as Monte Carlo Analysis to determine a statistical distribution of probable outcomes based on the constrained variability of component parts of the system. The founding principles are that the physical components are identically reproducible within a confined manufacturing tolerance and that identical components will react in an identical way when subjected to a given stimulus.

When simulating systems such as traffic, the component parts do not have the same confined reproducibility and are not constrained to react in a predetermined way to a given input opportunity. The response to any stimulus is therefore a probability function.

### Human Interaction and Stochastic Flow

Traffic flow may be considered to be stochastic based on the number of individuals involved all making independent decisions. The number of independent decisions that are possible in any traffic flow immediately identifies that there is a psychological element involved in traffic flow; if this was not the case then all drivers would make the same decision for every potential situation. Without being able to rely on absolute and repeatable physical constants, all vehicle drivers are forced to make decisions based on estimation of the current and intended immediate future location, atmospheric effects on the road surface, vehicle capability, own capability and an anticipated interaction with other nearby vehicle systems, animate and inanimate objects. The interaction with external animate and inanimate objects cannot be ignored, while a degree of influence and expectation can be applied to other vehicles, the same cannot be applied to meteorological, geological or animal based activity.

Beyond the vehicle control considerations, there is also route choice decision. Road systems that do not have alternative paths are liable to become gridlocked under certain conditions, for example vehicles can easily become stranded between junctions on motorways / dual carriageways if an accident causes the road to be closed. Conversely, where an alternative route exists, a probability exists that a percentage of traffic is likely to utilise the alternative; the probability of use increases in direct response to the apparent delay (congestion) on the preferred route. The probability will range from approximately zero, when the main route is observably clear and free flowing, to approximately one, when the main route is observably closed. The decision however is the responsibility of the driver and not the environment. Where the entire distance covered by the main and alternative route is not observable, the driver may also make a decision based on previous experience, observable conditions and external data such as satnav, TrafficMaster or public radio services. The decision may also be influenced by the drivers view of their relative importance or journey urgency as unilaterally compared to other drivers. This can be related directly back to Game Theory covered in section 2.4.6. Drivers are able to make and change route decisions at will and can only be forced to proceed in a continuous decision when there are physical barriers to prevent an alternative.

Section 3.4 briefly discussed the concept of dilemma zones at traffic signals, these being the physical and temporal area approaching a yellow or recently indicated red signal. Within this area, drivers will make different decisions whether to stop or proceed based on ambiguous and uncertain attributes such as current speed, traffic flow, importance / urgency of journey, likelihood of causing an accident and probability of prosecution. The number of drivers passing through the junction at any instant after the yellow indication is likely to be a Markov in nature, a normal distribution by count but not by order; only a probable prediction can be made regarding whether a vehicle will or will not pass the signal point.

Many of the papers written by and about Aimsun include reference to the (pseudo) stochastic features that are used [59]. The number and complexity of the features that are simulated within commercial traffic simulation packages is highly indicative of the influence that human response is perceived to have on traffic flow. This also suggests the level of investment that is made with regards to psychology research and the limitation of purely mathematic methods in traffic prediction.

### Limitations Of Individual Optimisations

There are many methods that can be utilised to analyse and optimise traffic flow, some of which have been summarised in this chapter. Each of the methods discussed has some apparent benefit however none can be considered as a complete solution. As an example, queueing theory is seemingly the most suited to the problem of identifying and resolving problems associated with traffic systems however it does not allow for the component parts of a queue (the individual vehicles) moving at individual rates and therefore does not consider that queued elements cannot be moved as a block entity. It seems commonly assumed that vehicles can move in block formation, this is the justification for vehicle platoons and reducing headway in driverless vehicles. Block movement is only possible when there is no acceleration component to consider e.g. electrical signals. Ignoring this limitation makes several incorrect assumptions:

* To move as a block, a number of vehicles would occupy the same roadspace regardless of speed, including stationary.
* Stationary vehicles must maintain a headway that is safe for the highest speed that they will achieve otherwise they must act individually to adjust inter-vehicle spacing.
* All vehicles in a stationary block must begin to move simultaneously in order to maintain the block spacing.
* The entire block will accelerate at the same rate as the slowest vehicle and decelerate over the same distance as the heaviest vehicle. This is regardless of the position of the slowest / heaviest within the block. Traffic movements will take longer, queues will form faster and congestion will increase.
* Traffic queues would not compress as they slow to a halt as this implies that they move as individuals rather than as a single block entity. The result is that stationary vehicles will occupy more space than necessary, queues would therefore occupy more rather than less space; congestion would increase for the same number of vehicles.

In order to avoid this problem Queueing Theory must be combined with e.g. Cellular Automata. In this case, traffic queues would advance according to the Cellular Automata rules for the duration of μ in the Queueing Theory. This is similar to the Petrinet concept. Using this approach causes the problem to be divided into the same number of parallel streams as the number of junction entry points. Each stream must be addressed simultaneously. A further complication is that of the individual driver / vehicle response and capability; this must be addressed by the inclusion of Human Interaction components.

While an understanding of the fundamental mathematical principles is highly desirable, it is not an absolute necessity when there are reliable traffic simulation tools available. The major concern is identifying which are reliable and able to simulate real scenarios to generate realistic results.

The traffic simulation tools are able to converge to a solution for complex networks in a relatively short period of time which is major benefit over manual application of mathematical principles.

It is clear from the stochastic and human response considerations that are employed in commercial traffic simulation packages that, although considered necessary in traffic simulation and prediction, discrete mathematic functions are not able to provide a complete solution. In use, it is apparent that Aimsun is reliant on prediction and simulation of human response. The main limitation of the simulation solution is the number of available variables that can be utilised, many of which are preconfigured but may be adjusted for any number of reasons.

Traffic simulation tools incorporate and automate the application of mathematical methods and combine elements of less defined concepts including Chaos Theory and Game Theory. Inclusion of probabilistic based assumption may help but cannot be relied upon to give definitive and irrefutable prediction of future events.

It is ultimately concluded that existing tools and methods are able to provide a reasonable approximation over a modest time period; absolute traffic prediction is not possible.

# Congestion Modelling

## Outline

The definition of congestion is very wide ranging, it can be more personal and emotive than it is scientific and empirical. It is often quoted that congestion occupies time and space which seems obvious, however the meaning can be much deeper. For example, a journey in a midmorning period could be considered to be “in congestion” even when the empirical factors of number of vehicles and length of time to complete the same route at an evening rush-hour are higher but only considered to be “slow moving”. The traffic speed in a major city considered as normal or light would almost certainly be thought of as congestion in a small town and, even more strangely, this can be the view of the same driver on different legs of the same journey or making the same journey at different times of the day. It seems that the definition changes with circumstance, moderate length traffic inner city delays would be considered gridlock if they were in a suburb, however that same level of traffic is somehow downgraded to only slow moving when it predictably repeats. There is also a difference between what constitutes congestion when observed in urban areas, point to point motorway and orbital motorways.

Within the context of this work and based on the initial findings, the definition of congestion will be based on the unintended number of stops made in a journey and the additional time taken to complete a journey. This definition is more closely related to the overall system waste and hence harmful by-products of transportation.

Congestion rates and locations for all road types change over periods of time. The changes can be cyclic, sudden step change or gradual drift due to change of destination

All road types can be subjected to congestion during extreme circumstances such as major accidents and incidents, when there is an urgent priority such as emergency services access or during mass attendance meetings such as sports events.

Congestion is not limited to motor vehicle traffic, congestion can be observed with cycles and is commonly observed in pedestrian traffic during peak times at underground, rail and bus stations as well as in shopping centres, supermarkets and at sports stadia. The important concept is that it is people (or goods) that are being moved and not the vehicles. Congestion is a direct consequence of movement within a restricted area, the area restriction is a consequence of the destination popularity.

## Congestion Effects

Congestion is normally viewed as a completely negative subject that should be eradicated at all costs. While there are indisputable negative aspects to congestion, there are also some aspects which make congestion desirable, if not essential.

### Negative Aspects of Congestion

The problems of urban traffic congestion are widely publicised. One of the most commonly observable effects is that the close proximity of vehicular traffic will directly result in a higher accident rate and potentially a higher injury rate. The discussion in section 1.8 (scope of work) and UK DfT Road Accident Statistics support the proposition of increased accident rates however, the relatively slow speed especially during congestion periods mean that KSI (killed or seriously injured) rates are much lower per accident.

Congestion directly affects everyone in the civilised world and many people in developing nations. Indirectly it affects the whole planet including uninhabited regions. While time delays due to congestion have been reported since Roman Times, the massive but indirect effects of congestion have only become globally significant since the advent of prime movers which rely on combustion processes to create motive power. It is worth noting that congestion is not limited to vehicular traffic; the subject of congestion in pedestrian traffic (without any vehicles) is widely studied to develop evacuation plans and crowd control processes [3] [5] [4]. Transportation has played a fundamental role in the developing world and will continue to do so. It is inevitable that new or alternative power sources will be developed and utilised in the future to maintain, among other things, transportation.

Aside from the obvious delay in being able to move from one place to another in a minimum and consistent time, there are reports that identify issues including physical and mental health, increased global warming and economic penalty due to inefficient fuel use, economic impacts due to insurance costs and increased accident rates due to vehicle proximity. Statistics are available that show a coincidence of higher numbers of road casualties at periods of peak congestion [60]. Inefficient fuel use is not limited to internal combustion vehicles; hybrid or electrically powered vehicles also waste energy when occupied.

Some of the globally significance concerns and effects of combustion derived power are:

* the rate of use of non-renewable resources including coal, gas and oil
* release of toxic compounds such as carbon monoxide (CO) and Nitrogen Oxides (NOx)
* release of environmentally damaging compounds including greenhouse gasses (GHG) such as carbon dioxide (CO2), nitrous oxide (NO2)and acid rain components such as sulphur dioxide (SO2)
* Locally, the release of heavy particulate matter (soot) is damaging to health and the built environment.

Alternative fuelled vehicles (AFV) are available but this does not resolve the immediate problems for several reasons:

* The power source for pure electric vehicles is likely to originate from fossil fuel power stations. In the UK during 2013 power derivation consisted of 70.4% coal/gas/oil, 17.6% nuclear and 12% renewable sources [6].
* Low uptake. The Society of Motor Manufacturers and Traders (SMMT) figures show 1.38 million new car registrations in Q1 and Q2 of 2015 of which less than 38 thousand were some form of AFV [7]. Conversely, over 97% of vehicles sold were traditional combustion powered.
* The majority of new electric AFV are hybrid electric / internal combustion rather than pure electric (23,500 v 14,000 vehicles in Q1 and Q2 of 2015) [7]. Therefore, over 62% of vehicles were combustion capable.
* Solar cells and batteries have finite lives and need ongoing replacement. Construction methods and materials are hazardous, disposal may be a developing problem.
* Electric vehicles use energy while occupied due to heating, cooling, lighting, entertainment etc. The energy is wasted if the vehicle is serving no useful purpose due to being stationary.

Manual transportation methods such as walking and cycling do not satisfy the needs of developed societies who require bulk deliveries of food and other goods into the city centres. Manual methods are not suited for less able-bodied, longer distance, substantial load, extreme weather conditions, professional business or events requiring fast response.

Public transport systems are limited by destinations and not well suited to lower density suburban or rural residential areas. In urban areas, road based public transport held in congestion may create greater environmental problems; prioritising a relatively low number of busses over cars serves to increase the delay for the higher numbers of cars. Ultimately due consideration needs to be given to the fact that the requirement for transportation is to move people and goods and not the vehicles that carry them.

There is no simple universal solution. Reducing congestion duration in urban areas has the most impact, is achievable in the shortest timeframe, does not require cultural shift and will continue to be effective in future systems.

### The Economic Benefit of Congestion

Economic viability of local, national and global commercial entities is not always negatively affected by congestion caused by increasing vehicle use. In some cases, queues are desirable and possibly essential.

The simplest and most obvious method of reducing congestion in any given area is to create a method to divert traffic away from the area. The effect of creating a (vehicle or pedestrian) bypass route reduces the amount of traffic but also reduces the advertising capability and number of potential customers to any retail outlet on the original route. In a small town, it is speculated this can have a devastating economic effect with a spiral of reducing retail profitability causing shops to close, further reducing retail relevance, reducing profitability and so on. For an outlying village environment, bypassing could increase property prices and be an economic benefit however it can also make the area less accessible with route priority given to the bypass traffic. The proximity of the bypass could cause an area to suffer many of the negative environmental effects of traffic flow but without direct access to the transportation links necessary for commuter traffic.

Stationary traffic can be advantageous for retail outlets as queues can maximise free shopfront advertising. Smaller shops may benefit from the relative convenience to passing traffic. Queueing for access can be seen as a psychological recommendation. Limited availability of an item or service can create an additional demand regardless of actual value or benefit. Examples include availability of seating at a restaurant, purchase of a newly released iPhone version or car parking at a shopping centre. Queueing at the shopping centre implies that there is something worth attending for, which is likely to attract further attention and generate further queues. Conversely, a lack of queues can imply a lack of interest or, from a retailer perspective, a potential for oversupply which may minimise profitability.

The economic success of a town is fundamentally concerned with the number of businesses and people who want to reside in the area. Each resident entity is subjected to a form of local taxation in the form of Business Rates or Community Charge. There is a symbiotic relationship between businesses and residents. The subject is beyond the scope of this thesis but the fundamental principle exists that people are attracted to areas with better employment prospects and businesses can proper in areas where there is a willing and capable workforce close to hand; businesses tend to be concentrated into small areas. The increasing prosperity of the region dictates that there will be an increasing need for housing, the housing requires greater land area and so there is the development of “urban sprawl” as a direct and unavoidable consequence. As the residential area expands, the travel distance to the employment areas increases and the need for transportation becomes more acute. The increasing transport leads to delays, queues and congestion. Congestion can therefore be a means to evaluate regional economic success.

It terms of transportation, slow and stationary traffic is more likely to permit courtesy response for less dominant entry to primary routes; fast moving vehicles are less likely to make a substantial speed reduction to allow another user to gain a position advantage. This is a fundamental aspect of game theory, the players should benefit from their actions. When allowing another vehicle to join a road, there is no perceived time loss to a driver in stationary traffic but there is psychological social, moral and ethical “good deed” reward. In moving traffic, the driver would need to lose time and psychologically disadvantage themselves for the benefit of another vehicle with no further benefit.

Congestion can be used as a means to confirm correct road sizing. The theory is that if a road does not experience any delay at any time then the capacity of the road exceeds the justifiable need and the roadway size could be reduced. The reciprocal argument is therefore that a road which experiences queueing for a minimal period during peak times is adequately sized to meet the design requirement. Clearly there is a minimum size requirement for access, but the access could be reduced to pedestrian traffic or even removed in extreme cases. The space dedicated to publicly accessible roads does not generate revenue for either real estate developers or regional councils and therefore, for economic reasons, the minimum area is used. The determination of the correct amount of queueing time is outside of the scope of this thesis.

A similar sizing argument can be applied to busses, trains, taxicabs and other driverless vehicles. The argument can explain why there should never be enough transportation service available to satisfy a peak need. As a further extension, for maximum efficiency, the level of transport (as a service) available should always be reduced to the current requirement level. This is typically achieved by reducing frequency of busses, reducing numbers of carriages on trains and the number of taxicabs in service. Other than privately owned vehicles (POV), transportation is rarely available at the point of need without forward planning. The combination of availability and location may result in a distributed queue for service dispersed over a wide area. A number of people waiting at a taxi rank is readily observable as a queue whereas the same number of people distributed across a number of homes or offices is only observable as a queue by the service provider.

On a wider economic scale, the level of vehicular congestion is directly related to the overall number of vehicles in use. The number of vehicles in use is directly related to the number of vehicles sold and to the service life of the vehicles. According to the Society of Motor Manufacturers and Traders (SMMT), the value of UK vehicle exports was £27billion in 2011 and 720,000 people were directly employed in automotive sectors [61], the figures increase to £34.3billion export (12% of UK total goods) and 814,000 employees in 2015 [62]. A subsequent effect of reducing vehicle congestion could be a reduction in vehicle production. The economic impact of reduced vehicle production is complex and may include diverse topics of corporate and personal taxation, costs of unemployment, effects on health, costs of production on a reducing scale, increasing costs of component parts in a reducing marketplace and so on. Clearly these topics are outside of the scope of this thesis however it is reasonable to assume that the impact at a national level would be undesirable.

### Challenges of Congestion

Congestion is a highly complex problem. The observable causes, effects and solutions are multi-facetted and highly superficial obscuring the underlying aspects of local, national and global economics, social and professional mobility and freedom of choice.

In terms of observable effects, congestion is not a purely physical entity which can be measured empirically; other major components of a congestion diagnosis include psychological and emotional impact. The notable effects of congestion, and therefore the influences that need to be resolved, depend on the motivation of the group investigating.

Some examples of problems to be resolved may be the length of queues, delay in completing a journey, delay in completing a journey section, amount of fuel being used, amount of excess fuel being used, levels of chemical pollution, levels of noise pollution, seismic events, loss of green spaces, and so on. The list is by no means complete and each item will normally have a temporal component, for example vehicle queues may only be excessive at certain times.

It is relatively easy to focus on a single problem and therefore to attempt to resolve only that one item with no benefit or possibly to the detriment of other issues. A very simple example of a single solution is the introduction of alternative fuel vehicles (AFV) such as stored electric energy (battery electric) propulsion systems. These vehicles may address local environmental damage but have no beneficial impact on vehicle queues or delays.

## Quantifying Congestion

As in all experiments, some form of measurement is required to be able to compare effects and a significant challenge is how to measure something that is not absolutely defined. Common measurements include number of vehicles that pass a point in a given time period, average speed of vehicles on a section of road, average speed for a “trip”, length of queues, number of stops, duration of stops, duration of “trip”, standard deviation from “normal” (count, speed, stops, duration of stops). The parameter chosen to provide the most relevant comparison in this work is delay per mile although other data has also been retained for completeness.

Within the subject area of this project there is tremendous scope for sensationalism and “headline grabbing” reports. For example, Forbes reported a story quoting 4.8 billion hours lost in the USA due to congestion in 2011, a cost of $101 Billion (US Dollars), a cost increase of 380% since 1982 [63]. The story is based on an annual report from the Texas A&M Transportation Institute [64]. While the figures are not disputed, the 2012 data from table “exhibit 1” gives an annual delay figure for 2011 of 38 hours per commuter, an annual fuel waste of 19 gallons and production of 380 lbs of CO2. Assuming the data is applied from a twice daily journey, 5 day week, 48 week working year, it is far less sensational to consider under 5 minutes delay and 0.04 gallons of excess fuel journey leg per day. Considering the context of the cost increase, according to the USA Social Security Administration, average wages in 2011 were 180% higher than in 1982 [65] and fuel prices were 170% higher [66] which is a very significant proportion of the cost increase. Of significance though is that according to the report [64], the number vehicles in the USA doubled between 1982 and 2011. The real costs excluding wages and fuel prices were generally of the same proportion which strongly implies that vehicle (hydrocarbon) fuel efficiency for has not significantly improved.

Inrix produce frequent reports which analyse statistics related to traffic congestion on a global scale, amongst other topics. The reports are available on the Inrix website [67].

### Road Capacity

An understanding of road capacity is fundamental in any transportation study. Road capacity establishes the number of vehicles that are able to traverse a section of road in any given period.

Accurate actual capacity and estimated capacity are fundamentally important components for road network design and traffic simulation. Minderhood [68] explains that there are at least three different capacity values namely Design, Strategic and Operational and states their importance in differing situations. He further explains that the estimation of capacity is a difficult engineering problem, is variable, probabilistic and that it has a stochastic element.

Van Rijn [69] uses data from the Highway Capacity Manual [70] [71] to explain “adjustment factors” that are applied to the “basic saturation flow” of 1800 PCU/hr. According to Ha et al [72], gross TH refers to the time between the same aspect of consecutive vehicles passing the point, commonly the leading surface of the vehicle. Nett TH or time gap refers to the time between the trailing edge of a vehicle and the leading edge of the following vehicle.

There are several alternative methods of calculating or measuring road capacity, including headway, occupancy and physical measurement. The following sections consider per lane road capacity.

#### Headway

Headway is used as one method to calculate road capacity. Headway is the time taken for the same aspect of consecutive vehicles to pass a fixed point, e.g. front bumper to front bumper. It is frequently quoted as 1800pcu/hr (passenger car units per hour). This is as quoted in many sources including [68], [69], [71] [70].

The following work was undertaken as a means to understand the derivation of this value.

For all cases:

Equ( ‑)

The capacity of a road in its simplest terms is the number of vehicles travelling at a given velocity with given headway could pass a fixed point in a measurement period. This is the same as:

Equ( ‑)

A common headway value used in the UK is 2 seconds, approximately the same distance suggested as being a safe separation gap to the vehicle in front. Substituting the headway period of 2 seconds and a measurement period of 3600 seconds (1 hour) into Equ( 3‑1) and Equ( 3‑2) :

Equ( ‑)

It can be seen that in Equ( 3‑3) all of the measurement units cancel and the final result is simply a value equivalent to the number of 2 second intervals in 1 hour. The number assumes that every vehicle occupies a 2 second window in the time domain irrespective of speed or vehicle size.

#### Occupancy

shows the physical measurement of timed headway and compares this to a value of gap and occupancy measurement. Headway is a relatively simple measurement to obtain by stationary roadside equipment however it is not practical for a following vehicle to determine the position of the front of the vehicle ahead and hence determine headway. In addition to this, at slow speeds the lead vehicle may take longer than 2 seconds to pass a fixed point and therefore the lead and following vehicle would occupy the same space concurrently. From the examples shown in Table 3.1, an average car of length 4.5metre travelling at 5mph would need more than 2 seconds to travel its own length; a bus would require more than 2 seconds at 13mph.

Table .1: Speed for Vehicles to Move Own Distance in 2 Seconds

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Vehicle | length | speed to travel length in 2 sec | | |
| m/s | mph | km/h |
| Car | 4.5 | 2.25 | 5.033093 | 8.099957 |
| Bus | 12 | 6 | 13.42158 | 21.59989 |
| Rigid lorry | 12 | 6 | 13.42158 | 21.59989 |
| Articulated lorry | 16.5 | 8.25 | 18.45467 | 29.69984 |

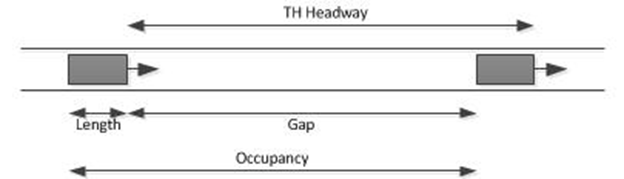


Figure .1: Occupancy vs Headway

Occupancy is the length of the vehicle plus the time gap to the vehicle in front. The gap to the vehicle in front is used because it is influenced by the vehicle of interest, the gap behind is controlled by the vehicle behind. It is not practical for a vehicle to control a trailing gap size.

Considering the vehicle length, the distance covered in the Occupancy period is given by:

Equ( ‑)

Assuming a 2 second gap rather than a 2 second headway, Equ( 3‑3) becomes:

Equ( ‑)

As the measurement units no longer cancel, Equ( 3‑5) has a dependency on speed and, equally significantly, a dependency on vehicle length. From this equation it can be seen that road capacity increases with speed.

Table .2: Comparison of Capacity Based on Headway and Occupancy

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| speed (mph) | distance (ft) | | Capacity | |
| hour | 2 sec | headway | occupancy |
| 10 | 52800 | 29.33333 | 1800 | 1197.357 |
| 20 | 105600 | 58.66667 | 1800 | 1438.096 |
| 30 | 158400 | 88 | 1800 | 1541.399 |
| 40 | 211200 | 117.3333 | 1800 | 1598.824 |
| 50 | 264000 | 146.6667 | 1800 | 1635.379 |
| 60 | 316800 | 176 | 1800 | 1660.693 |
| 70 | 369600 | 205.3333 | 1800 | 1679.259 |
| 80 | 422400 | 234.6667 | 1800 | 1693.458 |

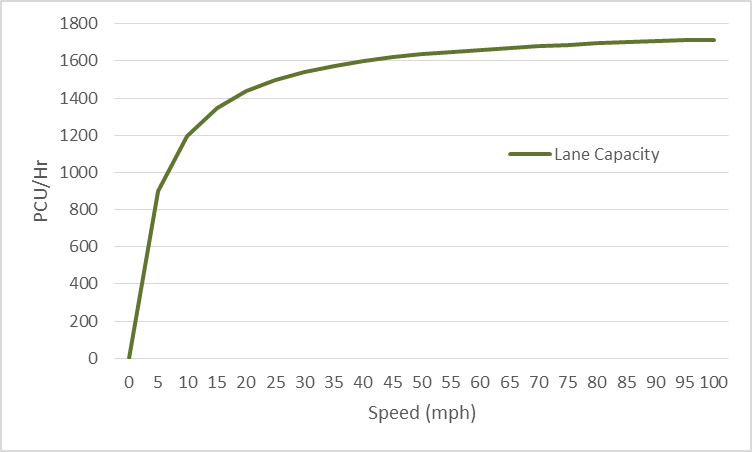


Figure .2: Theoretical Lane Capacity with 2 Second Gap

#### Measured values

Many papers have included actual measured capacity values and it is common for these values to apparently contradict the calculated values.

Oh and Yeo [73] reproduced figures obtained by other researchers that show values on multilane roadways as high as 2563 vehicles per hour although these rates were typically of short duration. The relevant figures are reproduced in Table 3.3

Table .3: Measured Capacity on Multilane Roads

|  |  |  |  |
| --- | --- | --- | --- |
| researcher | lanes | capacity | per lane capacity |
| Cassidy 1999 | 3 | 7000 | 2333 |
| 6890 | 2297 |
| 7120 | 2373 |
| 3 | 6490 | 2163 |
| 6620 | 2207 |
| 6120 | 2040 |
| Bertini 2004 | 2 | 4475 | 2238 |
| 4440 | 2220 |
| 4340 | 2170 |
| 4530 | 2265 |
| Bertini 2005 | 2 | 3690 | 1845 |
| 3690 | 1845 |
| 3840 | 1920 |
| 3750 | 1875 |
| 3510 | 1755 |
| Cassidy 2005 | 4 | 10250 | 2563 |
| 9250 | 2313 |
| 9500 | 2375 |
| 9730 | 2433 |
| 10100 | 2525 |
| 9600 | 2400 |
| Chung 2007 | 2 | 4070 | 2035 |
| 3 | 6500 | 2167 |
| Average |  |  | 2189 |

The feasible reason for actual capacity to exceed the theoretical value is that the gap between vehicles is much less than two seconds. Reasons for the gap to be below 2 seconds include combinations of psychological conditions of impatience, feeling safe with own ability, trust in vehicle protection measures, the need to convey “hurrying” to other road users or the inability to be able to measure a distance gap of two seconds between moving vehicles. The less feasible alternative is that the speed is orders of magnitude higher than the maximum speed used in Table 3.2. The concepts of impatience, hurrying and personal space within the context of London Underground travel are discussed in an article by Simeon Koole [74].

Measurements taken during headway studies have shown that vehicles commonly travel at distances much less than two seconds. Abtahi conducted a study and found distances of 0.24 to 5.98 seconds with medians of 1.18 to 1.5 seconds [75] and Moridpour reported distances for passenger cars of 0.09 to 4.9 seconds with a median 1.17 seconds [76]. It is extremely difficult for humans to accurately and repeatedly measure time spans without the assistance of some form of clock equipment and there is clear evidence that the perception of periodic time is inconsistent even within a single individual for reasons including emotional state [77].

### PCU (PCE) values

PCU or PCE is an abbreviation for “passenger car units” or “passenger car equivalents”. They are used as a measurement system which recognises that, for example, a bus, a passenger car and a bicycle do not occupy the same road space. It is highly desirable for traffic calculations to use a standard value for vehicle size. This simplifies calculation of capacity and flow given that they are affected by vehicle length. The simplification is obtained by PCU (Passenger car units) or PCE (passenger car equivalents). Significant work has been undertaken to calculate and utilise PCU / PCE values [78] [79], including at least one complete PhD thesis [80] .

Research has been conducted to determine exact “per lane, per direction, per hour” PCU values. Experiments have concluded ranges of 1600 to 2000 [81], 1700 [70] and 1800 to 2100 [82]. These figures are not an exhaustive list. A commonly cited value is 1800PCU/hr.

Using the figures from Table 3.1, i.e. car length 4.5m and bus length 12m, the information shown in Table 3.4 is constructed. Considering the road capacity in terms of bus traffic and comparing this to the capacity in terms of cars, the final column of PCE values is computed.

This is a very simplistic view of PCU which should be valid for free flowing traffic travelling at a constant speed over level ground. Real world traffic dynamics are complex and multivariate, they include transitions to and from stationary to maximum speed at varying rates; as a minimum the normal rate of acceleration and deceleration would need to be considered. The complexity of this topic is studied in detail in the PhD Thesis of Marlina [80] and in papers by Anand et al. [78] and Brooks [83].

The Highway Capacity Manual 2000 [71] lists a single bus as having a PCE 1.5 on level terrain, Anand concluded a value of 2.48 based on headway or 2.33 overall. Transport for London use a figure of 2.0 [84].

Table .4: Calculated PCU for Bus, Using Occupancy Value

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| speed (mph) | distance (ft) | | Capacity (occupancy) | bus PCU |
| hour | 2 sec |
| 10 | 52800 | 29.33333 | 768.5208 | 1.558003 |
| 20 | 105600 | 58.66667 | 1077.147 | 1.335097 |
| 30 | 158400 | 88 | 1243.62 | 1.239445 |
| 40 | 211200 | 117.3333 | 1347.769 | 1.186274 |
| 50 | 264000 | 146.6667 | 1419.074 | 1.152427 |
| 60 | 316800 | 176 | 1470.956 | 1.128988 |
| 70 | 369600 | 205.3333 | 1510.4 | 1.111797 |
| 80 | 422400 | 234.6667 | 1541.399 | 1.09865 |

### Traffic Density

The previous sections showed that there is a theoretical and mathematically proven vehicular road capacity. The capacity is measured in vehicles per time interval. There is a related quantity known as vehicle density which is measured in vehicles per distance. The abstract concept of vehicle density is very useful in considering traffic however the empirical values are useful only when the speed is also quoted.

Traffic density is very high for stationary vehicles, with each vehicle occupying only slightly more space than the vehicle length. Assuming a vehicle length if 14.75 ft (4.5m) and an inter-vehicle gap of 1.5ft (0.5m), the vehicle density is 325 vehicles per mile. In theory, as the speed increases, the physical distance between the vehicles also increases to maintain a safe time-distance to the preceding vehicle. The increasing gap means that there is lower occupancy per unit distance and hence lower density. Mathematically, the amount of vehicles on any given road section is higher at lower speeds which is an important concept for congestion; congestion is caused by traffic density however congestion increases traffic density. This explains to some degree why congestion builds rapidly but reduces slowly, a phenomenon that is very apparent in the “phantom traffic jam”.

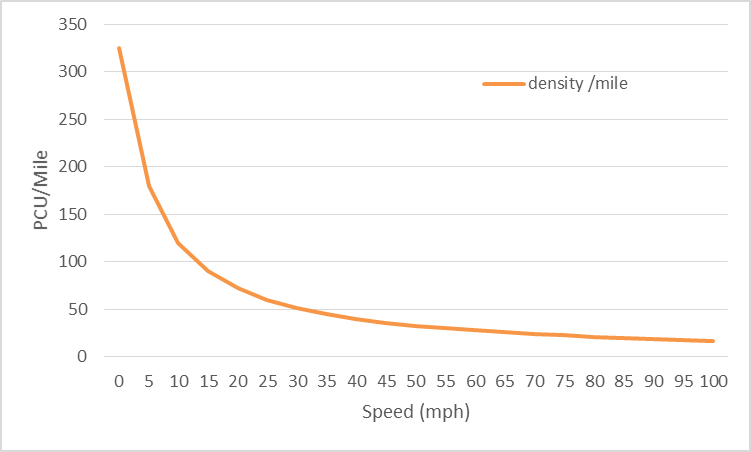


Figure .3: Vehicle Density per Mile At Various Speeds

In practice the physical distance does not always increase with speed. This is apparent in Table 3.3 where the measured capacity as far higher than calculated capacity. Rearranging Equ( 3‑5) and assuming a speed based on geographic location, the actual time gap can be determined for peak observed capacity.

Equ( ‑)

Equ( 3‑6) can be further simplified to

Equ( ‑)

Given that the samples are obtained on roads in the USA (except Bertini 2005), it is reasonable to assume a speed of 55mph. Using this value, a peak capacity of 2563 and assuming a vehicle length of 4.5m, the average gap is calculated to be 1.23 seconds. Each vehicle at this speed and length occupies (1.23s x 80.6fps) + 14.75 ft or approximately 114ft. This gives a vehicle density of 46 vehicles per mile as opposed to 30 vehicle per mile with a 2 second gap.

### Capacity Drop

Capacity drop is a phenomenon that has been observed whereby the per lane capacity of a road reduces after a restriction is removed. The capacity is therefore lower than anticipated a short distance after the road capacity increases. There are many studies which attempt to explain and / or quantify the capacity change for example Oh and Yeo [73] and Srivastava and Geroliminis [85].

### Braking Distance

In the UK there is an assumed safe distance between vehicles of 2 seconds. This distance is probably based on The Highway Code Rule 126 and Rule 168 [86].

Rule 126 states “leave enough space between you and the vehicle in front so that you can pull up safely if it suddenly slows down or stops. The safe rule is never to get closer than the overall stopping distance” and “allow at least a two-second gap between you and the vehicle in front on roads carrying faster-moving traffic and in tunnels”. Rule 168 concludes with “Drop back to maintain a two-second gap if someone overtakes and pulls into the gap in front of you.”

It is widely accepted that there is an average human response time of 1.5 seconds, although Mehmood and Easa found the average to be slightly less [87] and others have found it to be much higher.

Kamalanathsharma summarises several other works to provide deceleration rates in the range of approximately 6ft/s² (1.83m/s², 0.19g) to 13ft/s² (3.97m/s², 0.4g) [88]. The figures are based on vehicles arriving at a junction and are not maximum values. A paper by Gates, Noyce and Laracuente show figures as high as 20ft/s² (6.1m/s², 0.62g) and reports up to 25ft/s² (7.62m/s², 0.78g) [89]. Manufacturer data for automatic braking systems report maximum rates of up to 9.8m/s2 (32ft/s2, 1.0G)

Safe distance to a leading vehicle can be based on two distinct alternatives. The safe distance is either the time required for a following vehicle to react to the stimulus of a vector change of the leading vehicle or it is the entire distance required for a vehicle to stop.

In the first case, vector change, the assumption is made that the following vehicle can apply a vector change in the same timeframe as the lead vehicle. The two vehicles should, as a minimum, maintain a segregation which is proportionate to the difference between the initial time gap and the following vehicle reaction time. The vector change of the following vehicle could result in increased separation if the change in direction is different to the lead vehicle. In all cases, if the reaction time is greater than initial gap, there is a high probability of collision if the leading vehicle reduces speed.

In the second case an assumption is made that the vector change is neither indicated or anticipated, that only a portion of the lead vehicle makes the vector change (e.g. detachment or falling object), that the lead vehicle reveals a stationary obstacle that was previously obscured by the lead vehicle or that there is no alternative path and therefore no possibility to change direction. This situation could occur when there is an unanticipated object on the road e.g. heavy object falling from overhead or where traffic or other obstacles to the left and right allow only a single flow direction.

Table 3.5 lists the values given by the Highway Code [86] for thinking and stopping distances at various speeds.

Table 3.5: Braking Distances taken from The Highway Code

|  |  |  |  |
| --- | --- | --- | --- |
| Speed  (mph) | Distance (ft) | | |
| thinking | braking | total |
| 20 | 20 | 20 | 40 |
| 30 | 30 | 45 | 75 |
| 40 | 40 | 78 | 118 |
| 50 | 50 | 125 | 175 |
| 60 | 60 | 180 | 240 |
| 70 | 70 | 245 | 315 |

Dividing the “thinking distance” by the velocity in feet per second (fps) gives the calculated “reaction time” in Table 3.6. The time for braking is more complicated as it relies on the deceleration rate for a given vehicle which is in turn dependant on other factors such as road incline, friction between tyre and road surface, braking system and vehicle weight.

Given that the terminal speed, vterminal and initial time, tinitial, are both zero and that

Equ( 3‑8)

Equ( 3‑9)

Equ( 3‑10)

Equ( 3‑11)

Equ( 3‑12)

Equ( 3‑13)

Equ( 3‑14)

Equ( 3‑15)

Substituting the highest noted deceleration value of 25ft/sec2 (7.62m/s², 0.78G) completes the information for Table 3.6. Of note is that at any speed above 30mph, 2 seconds does not provide a safe stopping distance even with zero reaction time. These figures are especially significant when considering the potential for driverless vehicles to reduce congestion.

Table 3.6: Stopping Time Based on Highway Code Values

|  |  |  |  |
| --- | --- | --- | --- |
| Speed  (mph) | Calculated times (sec) | | |
| reaction | braking | total |
| 20 | 1.47 | 1.26 | 2.73 |
| 30 | 1.47 | 1.90 | 3.36 |
| 40 | 1.47 | 2.50 | 3.96 |
| 50 | 1.47 | 3.16 | 4.63 |
| 60 | 1.47 | 3.79 | 5.26 |
| 70 | 1.47 | 4.43 | 5.89 |

For comparison, stopping times based on initial speed and a maximum deceleration rate of 1G are given in Table 3.7. Based on the preferred rates obtained in [88] and [89], a continual or repetitive rate of 1G is highly undesirable. Again ignoring reaction time, 2 seconds does not provide an adequate distance at any speed above approximately 45mph.

Table .7: Stopping Times Based on 1G Deceleration

|  |  |  |  |
| --- | --- | --- | --- |
| Speed  (mph) | Calculated times (sec) | | |
| reaction | braking | total |
| 20 | 1.47 | 0.91 | 2.38 |
| 30 | 1.47 | 1.37 | 2.84 |
| 40 | 1.47 | 1.82 | 3.29 |
| 50 | 1.47 | 2.28 | 3.75 |
| 60 | 1.47 | 2.73 | 4.20 |
| 70 | 1.47 | 3.19 | 4.66 |

### Fundamental Diagrams

Graphs, known as Fundamental Diagrams, can be drawn to show the relationships between traffic speed, flow and density. The diagrams are frequently used as proof that, amongst other things, traffic flow will reduce once a critical density has been reached. The diagrams are originally attributed to Greenshields in a paper from 1935 [90] however there are many updates and interpretations that have been created by different researchers e.g. Smulders (1990) [91] and del Castillo (2012) [92].

The three diagrams are speed / flow, speed / density and density / flow. In the diagrams u refers to speed, k to density and q to flow. Subscript 0 is the initial state, c is the critical value and j is the jam value.

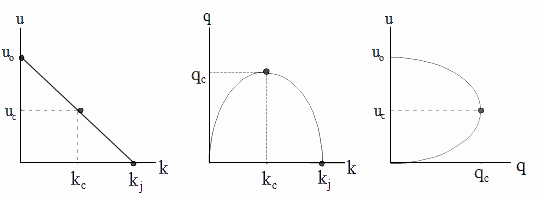


Figure .4: Fundamental Diagrams after Greenshields [90]

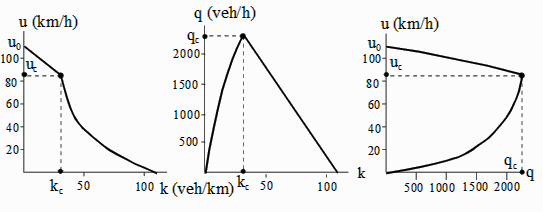


Figure .5: Fundamental Diagrams after Smulders [91]

The significance of the critical value and the relationship to congested flow is clarified in the density / flow diagram from Del Castillo shown in Figure 3.6. The diagram uses ρ in place of k to indicate density.

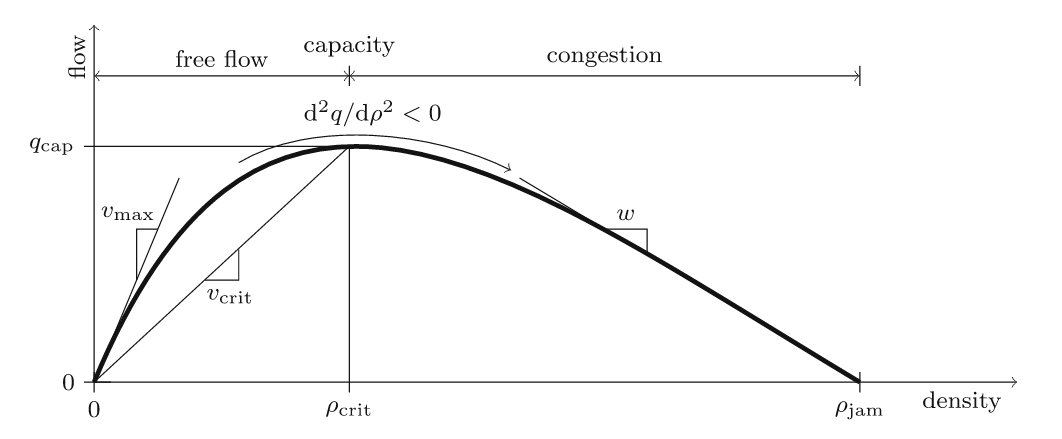


Figure .6: Density/Flow from del Castillo Relating Flow State to Critical Value [92]

Formal derivations of the fundamental diagrams are explained in a 2009 paper by Helbing [93] and an explanation of the historic development of traffic flow models is provided by van Wageningen-Kessels et al. [94].

Fundamental diagrams have greatest significance to inter-urban traffic but can be applied to urban traffic flow. They appear to have minimal significance to traffic flow through controlled junctions however one very important aspect is the explanation that all road sections have flow /capacity restrictions. Given this information, it is clear that it is not possible for a junction exit to have infinite sink capacity, even when there are no obvious downstream restrictions.

## How Isolated Traffic Lights Cause Delays

As highlighted previously, research has returned different values for PCU. As a baseline and to avoid misinterpretation, this section considers a traffic flow containing only passenger cars with a PCU of 1.0. A value of 1800 PCU/hr is used for the road capacity and for the flow rate. The traffic light controller is fully timed without any form of detection.

Modelling traffic at a junction is complex. By definition, traffic will arrive from multiple directions and attempt to utilise the same junction area. The following models are simplified for clarity in order to explain concepts. It must be recognised that at a simple crossroad junction, traffic will arrive from between two and four directions simultaneously with a corresponding number of queues.

Consider a straight single lane, one-way road with no traffic controls, as shown in Figure 3.7.

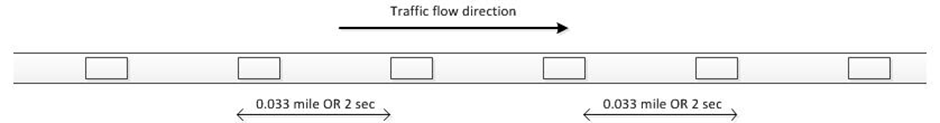


Figure .7: Single, One-Way, 30 Mph Lane at Full Capacity of 1800pcu/hr

Assuming a uniform distribution and a standard value of 1800PCU/hr it is calculated that, on average, 1800/60 PCU (vehicles) will pass a given observation point per minute. Alternatively this can be stated as 1/30 \*60 seconds per vehicle or 1 vehicle every 2 seconds, as stated in section 3.3.1.

As previously stated, occupancy and traffic density are related to speed. If the vehicles are travelling at a uniform speed of 30mph, the distance between leading edges (or trailing edges) for each vehicle is 2×30/3600 = 0.033 miles, the traffic density = 1/0.033 = 30 vehicles per mile.

Now consider adding a similar road as a cross junction controlled by traffic lights as depicted in Figure 3.8. Included in the simplification, no turns are allowed. This means that each direction is totally conservative; there is no interaction between the traffic heading from West to East (WE) and from North to South (NS). The traffic signals are ideal, showing either stop or go with no transition time.

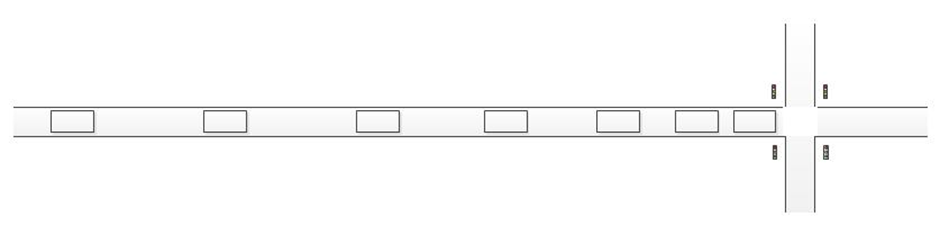


Figure .8: Single Lane with Controlled Crossroad

If the lights have a 60 second cycle time and 50% green duration in each direction then it is clear that the EW road direction is only “open” for 30 seconds of each minute. Using the simplistic model, 15 vehicles will be able to pass in any minute and therefore the capacity of the road has been halved. Directly related but alternative observations include:

* the roadway leading to the lights will be congested for 30 seconds of every minute
* the roadway away from the lights may be empty for 30 seconds of every minute
* the capacity in each direction is halved, therefore the total roadway capacity is also halved
* very importantly, the capacity of the actual junction is unchanged in this simple model

Figure 3.9 shows the theoretical equal 50% split, Head 1 provides WE flow control and Head 2 provides NS control.

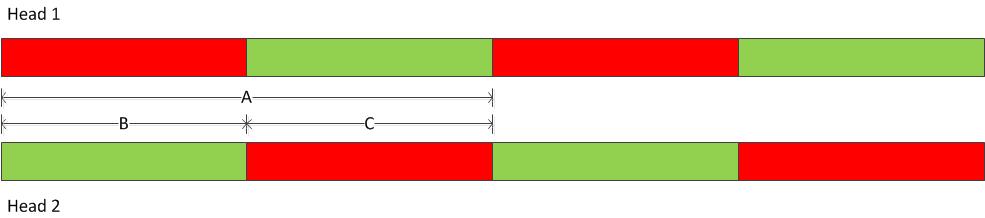


Figure .9: Theoretical 50/50 split

In the diagram, the time values for red and green on each head are identical, A represents the entire cycle, B represents red for signal head 1 (green for head 2) and C represents green for head 1 (red for head 2). The duration is denoted in the usual manner, using t with a subscript defining the event subject.

The cycle time is the time taken for a controller to display every light sequence and return to the initial condition. In the case of the controller in Figure 3.9:

Equ( ‑)

As previously stated, split is the amount of green time expressed as a percentage of the entire signal cycle.

Equ( ‑)

or, alternatively,

Equ( ‑)

which, if B and C are equal, simplifies to

Equ( ‑)

In a practical system, an equal allocation of green time between two opposing flows results in significantly less than a 50% split; traffic will flow for significantly less than 30 seconds. Practical signals in the UK include three additional stages; yellow (to advise that lights are changing to red), all red (to allow the junction to clear before a contending flow begins) and red/yellow (to advise that lights are changing to green. There are mandatory timings for the red/yellow (2 seconds) and yellow (3 seconds) [95]. The all red phase is an optional extension of between 0 and 25 seconds. There are several permutations possible that will incorporate the additional sequences; the method used in this paper is illustrated by Figure 3.10.

Important aspects of the sequence are that there cannot be any period where a go signal is shown to conflicting flows and that there must be a time period which allows the junction to clear before a conflicting flow begins. These requirements can only be satisfied by incorporating the additional stop periods into the natural green periods, therefore reducing the relative proportion of green time.



Figure .10: Practical Light Sequence

Referring to Figure 3.10, D is the cycle time, the measurement includes all light phases regardless of the initial point. The value of A from Figure 3.9 and D from Figure 3.10 is identical; the cycle time is unchanged.

E is the actual green time, while the effective green time is depicted by Fn,m where n is the head number and m is the sequence number. Due to the inclusion of human factors, the value Fn,m is variable and less accurately quantified as it includes part of the yellow time but not all of the green time. The effective green time starts after the actual green due to driver and vehicle reaction times. Studies have suggested that the delay is in the order of 1-2 seconds typically. The effective green extends beyond the actual green due to the yellow phase. The yellow phase is an instruction to stop if considered safe [96]. It is not physically possible for a vehicle to stop for a yellow signal if it is closer to the stop line than the maximum deceleration rate of the vehicle permits; it is not desirable to reduce speed at a rate which would alarm other road users or impact reasonable driver and passenger safety and comfort. It is expected that a vehicle will stop if the light changes to yellow in advance of a reasonable deceleration rate distance. These points give rise to the dilemma period, G, (or dilemma distance in the distance domain). Dilemma provides a time (or distance) range for a vehicle to stop rather than exact time.

The value of green time, E, is related to the value of green time C from Figure 3.9, by

Equ( ‑)

The actual red time, H, is given by the combination of all-red, red and red/yellow times. The effective red time, Jn,m, includes a portion of the yellow and a portion of the green as discussed earlier. The value of H is related to the previous value B by

Equ( ‑)

In terms of and , the cycle time is given by

Equ( ‑)

The entire sequence for Head 1 can be considered as F1,1, J1,1, F1,2, J1,2, …, F1,x, J1,x where x is the number of flow directions in the sequence. Note that the values of J1,m and F1,m are imprecise throughout the period of interest. The sequence for Head 2 is similar but described as J2,m and F2,m

If an assumption is made that the delay responding to a change to green and the delay responding to a change to yellow are approximately equal then . Using this assumption, the effective green time, , is reduced however the effective red time, , is increased.

In this case , split is determined as follows:

Equ( ‑)

Substituting Equ( 3‑22) followed by Equ( 3‑20) and Equ( 3‑21) into Equ( 3‑23), we obtain

Equ( ‑)

The actual utilisation time for the contended area of the junction is the sum of the split time for Head 1 and the split time for Head 2

Using mandatory 2 second red/yellow and 3 second yellow times, cycle times of 30, 60, 90 and 120 seconds with equal flow allocation in each direction, Table 3.8 and Table 3.9 are generated with “all red” times of 0 and 2 seconds respectively. These tables show the split time and the maximum possible junction utilisation.

Table .8: Split and Availability - 0 sec all red time

|  |  |  |  |
| --- | --- | --- | --- |
| Cycle Time  (sec) | all red time | split % | junction % |
| 30 | 0 | 33.3 | 66.7 |
| 60 | 0 | 41.7 | 83.3 |
| 90 | 0 | 44.4 | 88.9 |
| 120 | 0 | 45.8 | 91.7 |

Table .9: Split and Availability - 2 sec all red time

|  |  |  |  |
| --- | --- | --- | --- |
| Cycle Time  (sec) | all red time | split % | junction % |
| 30 | 2 | 26.7 | 53.3 |
| 60 | 2 | 38.3 | 76.7 |
| 90 | 2 | 42.2 | 84.4 |
| 120 | 2 | 44.2 | 88.3 |

Whilst the values used in Equ( 3‑24) could be adjusted to provide a 50% split time, this can only apply in a single direction and for a specific cycle time. As an example, to generate a 50% split for Head 1, *tE=tH+tY*or, given that is defined by Equ( 3‑20) and is defined by Equ( 3‑21), Equ( 3‑20)=Equ( 3‑21)+tY.

Equ( ‑)

Equ( ‑)

Using the known and values and assuming a zero value for :

Equ( ‑)

From Equ( 3‑16) and observing that ,

Equ( ‑)

Substituting Equ( 3‑28) into Equ( 3‑24) provides the data for Table 3.10. Split (1) % refers to the flow direction controlled by Head 1 while split (2) % is for Head 2. The table shows that while the junction utilisation is constant for a given cycle time, the opposing direction split is reduced as a result.

Table .10: Opposing Split and Junction Availability

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cycle Time  (sec) | allred time | split (1) % | split (2) % | junction % |
| 30 | 0 | 50 | 16.7 | 66.7 |
| 60 | 0 | 50 | 33.3 | 83.3 |
| 90 | 0 | 50 | 38.9 | 88.9 |
| 120 | 0 | 50 | 41.7 | 91.7 |

The “lost time”, or the period when no vehicles are passing, approximates to the sum of red/yellow, yellow and all-red times. This is 5 seconds (plus any all-red time) in every cycle as shown by Figure 3.10. Using the arrival rate of 1 vehicle every 2 seconds, every cycle loses a theoretical 2.5 vehicles (rounding to a practical value of 43 vehicles) per cycle.

For a typical 90 second cycle time with equal allocation and no additional all-red time, 22.5 vehicles could pass in each cycle. The loss reduces this figure to 20 vehicles. There are 40 cycles of 90 seconds in every hour and therefore 800 pcu /hour can pass in any given direction. This agrees with the split calculation of approximately 44.4%. Other values are shown in Table 3.11

Table .11: flow rates based on cycle times

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Cycle Time  (sec) | pcu/ direction /cycle | loss/ cycle | max pcu/ cycle | cycle/ hour | max pcu/ hour |
| 30 | 7.5 | 2.5 | 5.0 | 120.0 | 600 |
| 60 | 15 | 2.5 | 12.5 | 60.0 | 750 |
| 90 | 22.5 | 2.5 | 20.0 | 40.0 | 800 |
| 120 | 30 | 2.5 | 27.5 | 30.0 | 825 |
| 150 | 37.5 | 2.5 | 35.0 | 24.0 | 840 |

### Delay Periods at Traffic Lights

The previous section suggests that the best solution for traffic flow is to extend the cycle time as this permits the greatest flow rate however applying this logic does not result in the greatest flow rate.

Firstly, if the time period of Table 3.11 is extended, the chart shown in Figure 3.11 can be obtained. This clearly shows that while the number of pcu / cycle increases linearly, the maximum pcu / hour increases in line with an inverse exponential decay curve. Cycle times above 120 seconds are not recommended in the UK [97], this figure corresponds to a high level in the knee point of Figure 3.11.

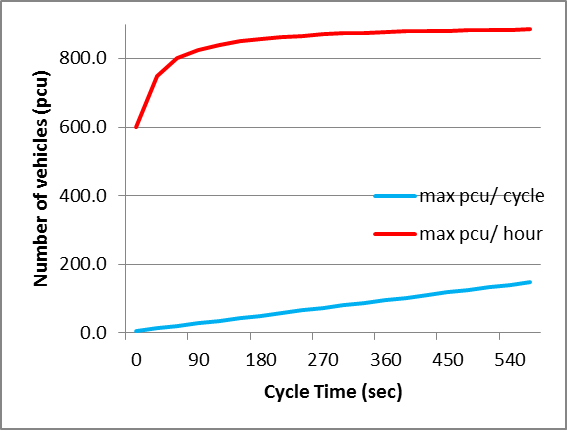


Figure .11: Effect of Increasing Flow Rate; Pcu Flow as a Function of Cycle Time

Secondly, the flow times are per direction. Increasing the cycle time gives greater flow time, and hence flow, in one direction but blocks flow for a greater period in the opposing direction. This is a two sided argument which may be viewed as psychologically negative, but it can be economically and environmentally beneficial if “auto-start” functions are available in all queued vehicles. It should be clear that for a constant vehicle arrival rate, the queue length is directly proportional to the block time, increasing the block time will require greater space for input queues.

Third, the theoretical benefit of increased flow is only realised if the flow rate is actually required. Long cycle times are unresponsive to low traffic flows. Traffic approaching a control point is in no way synchronised to the controller and may arrive at any time with entirely equal probability. If the signal is red on approach, the vehicle will have an average waiting time equivalent to 0.5 times the red period. The response to low volume dynamic arrivals is therefore much better with short cycle times.

The dynamics of a traffic queue are different for different traffic flow profiles. Where the maximum capacity of the road is in constant use (e.g. at or near capacity), fewer transitions are better as they represent less loss in a given period.

Where the road is operating below its maximum physical capacity, more frequent changes are better as they reduce the waiting time for vehicles when the opposing flow is not being used. Changes demanded in response to traffic detection offer significant benefit.

If the road is operating above capacity queues will not be able to clear within a cycle and the choice of cycle time appears largely irrelevant however longer cycle times should reduce the overall duration of the congestion period. If the junction is non-isolated, the longer cycle time can cause upstream blocking during the extended inactive (red) time.

Traffic arriving at a traffic light has a completely random profile; there is an equal probability of arrival at any time. Traffic arriving during the green period will pass straight through while traffic arriving during a red phase will have an average delay of one half of the red period. Using the 90 second cycle time from Table 3.8, the delay will be in the approximate range 0 – 50 seconds. Considering each second to be a discrete arrival period, the average delay is 25 seconds. The overall average delay when considering a vehicle arriving at any time in the cycle is actually 14 seconds because the delay range remains as the range 0 – 50 seconds but the arrival periods increased from 50 to 90. This apparently minor note is an example of where classic mathematical methods are unable to determine an absolute state at any given future time.

### Micro, Macro and Mesoscopic Flow

Microscopic flow is important for determining the flow detail for an individual vehicle. This can be used to calculate traffic flow such as exact numbers of vehicles and number of vehicles making a specific turn. While this is beneficial in some cases, it has a limitation of being unable to determine vehicle density and therefore does not present a general view of flow or congestion over time.

Macroscopic flow is a rolling value averaged over a period of time, considering the traffic to be a single contiguous entity rather than a collection of individual components. It does not contain the detail of a specific vehicle but represents a gross flow characteristic. Macroscopic flows are related to road use over time and can be useful for environmental studies and civil engineering planning. Visual observation of macroscopic flows often suggest the appearance of wave motion in fluid flow, the subject of the inspirational Lighthill-Whitham paper discussing congestion in 1955 [1].

Mesoscopic flow includes both flow averaged and specific individual vehicle data. Unlike macroscopic, vehicle interaction information is collected however it is not as complete as for microscopic. Mesoscopic measurement is used for combined vehicle and road use over time. Traffic models and simulation may be considered as mesoscopic if they seamlessly change between microscopic and macroscopic models (and vice versa) in response to optimising changes of physio-temporal regions.

The apparent information conveyed by micro and macro can appear contradictory. Consider a junction depicted in Table 3.11 operating with a 90 second cycle time and ignoring the vehicle loss due to traffic light transitions. If there is no initial queue and traffic begins to arrive at the junction at a rate of 1200 pcu/hr for a period of 30 minutes reducing to 300 pcu/hr for the subsequent 30 minutes, a queue will form and disperse. In each cycle:

Equ( ‑)

Where *Qdelta* is the change in queue length, *vehoff* is the number of vehicles attempting to use the junction in pcu/hr, *vehcap* is the junction capacity and *tcyc* is the cycle time. Using the values given, the queue will increase for the first 30 minutes/20 cycles at a rate of 10 pcu/cycle to a maximum of 200 vehicles. The queue will then decrease at a rate of 12.5 pcu/cycle until the queue is cleared which will take 16 cycles (24 minutes). A queue will be present for a total of 54 minutes. Using macroscopic measurements, the flow rate will be reducing for 30 minutes and then increasing for 24 minutes, returning to normal flow after 54 minutes. Macroscopic derived conditions show that there will be delays / congestion for at least 54 minutes.

For vehicles within the queue, progression is at the traffic signal service rate. While the overall queue length for a 90 second cycle is increasing for the first 30 minutes, the queue from the perspective of an individual vehicle is always reducing by 20 pcu/cycle (13.3 pcu/min). If a vehicle were to join the queue at the end of the 5th cycle it would be 51st in the queue. Using an average pcu of 5m including gap, the distance to the traffic light would be 250m. After 2 cycles (3 minutes), the distance to the lights would be reduced to 10 vehicles (50m). Using the values obtained in Table 3.8, vehicles move in the first 44.4% of the cycle (40 seconds), confirming the 2 second/pcu from Figure 3.7. The subject vehicle will arrive at the stop line after 20 seconds of the 3rd cycle. The total delay experienced is 3 minutes and 20 seconds. The maximum delay time for any given vehicle would be to join the queue at the end of the 20th cycle when the queue would be 200 vehicles or 1km. The delay would be 10 complete cycles or 15 minutes which is still far less than the macroscopic view of a 54 minute queue. Microscopic derivation shows that there will be a maximum delay of 15 minutes.

The previous calculations serve to clarify the difference in perspective between microscopic and macroscopic analysis of a queue. If the simplified macroscopic calculation of queue development is tabulated and the results plotted in a graph, the results shown in Figure 3.12 are produced. The figure shows that a greater number of cycles is required to resolve the queue if the cycles are shorter duration. The number shown in brackets in the legend indicates the cycle time. Something that is not immediately apparent from the graph is that product of cycle x period is identical for each line, the maximum queue length is identical and occurs after the same period of time. The figures are shown in Table 3.12. This contradicts the information shown in Table 3.11, which states that there should be a greater rate of clearance for longer cycle times.

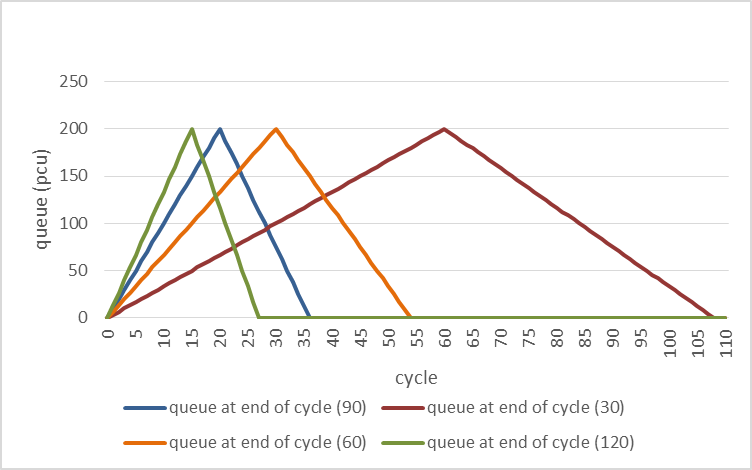


Figure .12: Queue length vs Cycles, Ignoring Transition Loss

Table .12: Queue Statistics for Zero Transition Losses

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cycle length (sec) | 30 sec | 60 sec | 90 sec | 120 sec |
| peak queue length (vehs) | 200 | 200 | 200 | 200 |
| time to peak (min) | 30 | 30 | 30 | 30 |
| cycles to peak | 60 | 30 | 20 | 15 |
| time to clear queue (min) | 54.5 | 54 | 54 | 56 |
| cycles to clear | 109 | 54 | 36 | 28 |

The discrepancy is caused by, and highlights the problem of, simplifying the calculation by ignoring the losses due to light transitions. If the previously calculated loss rate of 2.5 pcu/cycle is included, the graph changes to Figure 3.13. The detail events are shown in Table 3.13.

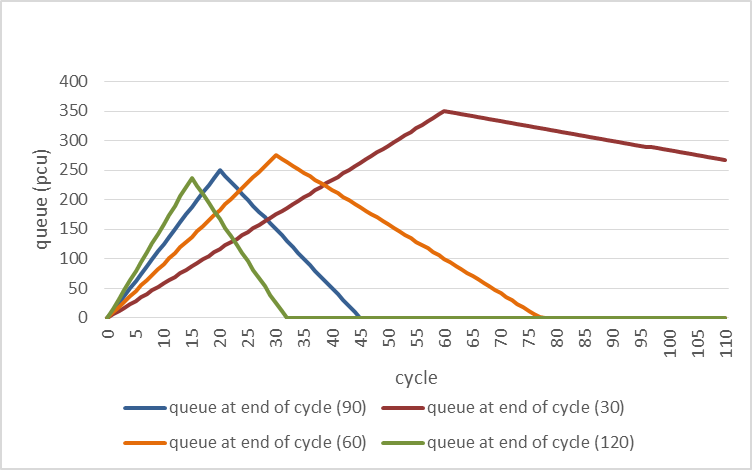


Figure .13: Queue length vs cycles, Including Transition Loss

Table .13: Queue Statistics with 2.5 PCU Transition Losses

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cycle length (sec) | 30 sec | 60 sec | 90 sec | 120 sec |
| peak queue length (vehs) | 350 | 275 | 250 | 237.5 |
| time to peak  (min) | 30 | 30 | 30 | 30 |
| cycles to peak | 60 | 30 | 20 | 15 |
| time to clear queue (min) | 135 | 78 | 67.5 | 64 |
| cycles to clear | 270 | 78 | 45 | 32 |

While the content of this section is intended to relate to simulation and modelling, the models also emphasise that highly dynamic traffic signals with shorter cycle times do not promote the best flow rates. This may be counterintuitive in the initial observation however this is easily explained by considering that there will be a loss at each transition. As the number of transitions increases, the time between transitions reduces and the percentage of cumulative time lost to transitions becomes dominant. When this happens, there will be a higher proportion of the cycle time given to loss rather than flow. The transition loss time is a function of the junction layout plus a safety margin, even when the safety margin is reduced to near zero, the junction layout will still dictate a loss period. The safety margin can never be reduced to absolute zero as this implies that vehicles must be in contact with each other for a significant period of time while they occupy the junction area. Conversly, rapid cycle times reduce the waiting time at a stop indication; where the offered traffic flow is light this can be beneficial.

### Causes of Congestion

It is clear that congestion is not omni-present; it has both physical and temporal start and end points.

Some event, at some point in time, causes congestion to start at a given location. Ultimately some other factor then subsequently either causes or permits the congestion to end. A common assertion is that simply the number of vehicles exceeding the capacity of a road (frequently quoted on traffic reports as “sheer weight of traffic”) but there are other causes and reasons why congestion may develop and continue. The capacity argument is vital to the discussion but it should be clear that the capacity of a junction or any road within the proximity of a junction (controlled or otherwise) is lower than the capacity of the same road in other areas. It should also be clear from the previous sections of this work that the over-capacity can be a fraction of a vehicle.

Beyond physical properties, some reasons for congestion can be political, economic, social or a combination of these and other factors. For example, slow or stationary traffic passing a government building could be considered to be a security threat; generally this type of building is located in busy cities. To reduce the threat, vehicles need to have minimal time near to the building preferably without stopping, the traffic speed therefore needs to increase and to achieve this, the volume of traffic must be reduced. A simple method to achieve this goal is to force traffic to queue at entrance points to the area. This is known as “gating”, a method of actively moving traffic density, congestion and queueing away from a nominated area however this results in congestion build up in the surrounding area. As another example, the popularity of a “place” (shopping centre, restaurant, theme park, city centre) can be implied from the number of people who are inclined to queue to be able to enter. The action of a queue can lead to the development of “FOMOphobia” (the Fear Of Missing Out), which leads to a compelling drive to own an item or be present at an event or location even without clear expectation of any outcome. Many works dating back as far as the late 1990s have shown that “the queue” can be very powerful advertising [98]. From these two simple examples it is clear that congestion can be considered a problem, a necessity or even a solution.

Traffic controls promote congestion even if they are not specifically the cause. The traffic controls may be intended as in the case of road markings and traffic signals or unintended such as parked cars, road works and accidents.

Specifically, traffic controls cause upstream congestion and relieve downstream congestion. Traffic light control systems are intended to control traffic flow in a predictable and repeatable manner. Existing traffic light systems detect, measure and control traffic flowing into an area and therefore actively create upstream queues. While it is not always the primary goal, this is the exact same action as “gating” described earlier.

### Congestion Taxonomy

Congestion begins when a vehicle approaches a point location in a time period at a rate which exceeds the flow capability of the point. This vehicle must therefore wait for service for a period which exceeds the regular service interval. The wait is the initial congestion queue formation. If the arrival rate continues at a constant level above the point flow capability, a queue will form. If the arrival rate increases, the queue will grow at an apparently linear rate. If however the rate reduces to a critical value equal to the point capacity, the queue length remains constant and does not dissipate. The arrival rate must fall below the critical level for the queue to clear, hence the hysteresis loop.

It should be clear that, using the above definition, a single vehicle difference in arrival rate can be the trigger for congestion to begin. However, small changes in the service time interval and slight differences in driver response can also be triggers for congestion. Vehicle Actuated traffic control systems vary the service interval. Variation in the service interval changes number of transitions in a period affecting the vehicle loss rate and therefore the average road capacity.

For a continuous arrival rate which exceeds the flow capacity of a road section by a fixed amount, a simple assumption would be that the congestion queue length will increase linearly. However the rate at which a vehicle is serviced in the queue is dependent on the distance from the service point, the time for the vehicle in front to be serviced and a reaction delay. The distance from the service point and the reaction delays are cumulative through the queue and therefore the service rate reduces with increasing queue length. The overall result is the queue length increases in direct proportion to itself. The service rate reduces in proportion to the queue length and therefore the effect of queue length on each consecutive vehicle increases. The queue experiences an exponential, rather than linear, growth rate.

An alternative view is found by considering an arrival rate at the critical level; natural variation in individual driver reaction time will eventually give rise to a reaction delay slightly higher than normal and cause the beginning of a congestion queue. The additional vehicle in the queue will slightly reduce the service rate, the original (and consistent) arrival rate will become increasingly marginal until the beginning of a congestion queue becomes inevitable.

A final observation is that the congestion queue actually begins some distance in advance of the restriction point, behind the vehicles that are part of the normal service queue length. At a traffic light controlled junction, the service queue length is directly related to the active (green) time period. VA controllers reduce the active period in response to vehicle arrival at conflicting entrances and may therefore inadvertently increase the likelihood and early onset of congestion queues.

The growth of the queue is dependent on arrival rate, which is a vector product with quantity, direction and time components.

As discussed in section 3.4.2, the time required for an individual vehicle to pass through a queue is not directly related to the queue dimensions; it is definitely not related to the duration of the queue. Queues therefore begin as a single vehicle at a single point. Similarly, congestion begins from a single point and spreads backwards from that point towards the approaching traffic. By this definition, congestion does not begin at the head of a queue but at a critical time and distance upstream of the restriction and is totally dependent on rates of arrival and dissipation rather than simply numbers of vehicles.

### General Congestion Resolution

General congestion implies the overall integral of all congestion at all periods. This is commonly used to identify emotive references to such items as “the cost of congestion to commuters” where even reputable sources publish articles such as “Traffic Congestion to Cost the UK Economy More Than £300 Billion Over the Next 16 Years” [99]. Suggested ways to reduce general levels of congestion frequently focus on one of several methods, some of which are considered below. It can be seen from the relatively short discussions that each has positive and negative aspects and that none are capable of satisfying all circumstances. At the ultimate stage, the entity that utilises the transportation medium must see some benefit of the choice made over the alternatives offered. The choice can be based on visibly apparent “obvious” choice e.g. travel distance or load to be transported, but may also be influenced by a number of other factors including cost, availability, political influence, environmental views, health consideration, physical ability and so on.

#### Remove Vehicles, Utilise Walking and Cycling

There is a view that all forms of un-powered transport are in some way totally cooperative and non-conflicting. Work on crowd control [4] proves that this is not the case even with a single mode transit system. With few exceptions they are relatively low payload, local use and may be socially exclusive to large sections of society including the young, old, infirm and disabled. In terms of carbon footprint, walking has minimal environmental impact however mass produced cycles have a carbon footprint based on manufacture. Modern carbon fibre cycles and power assisted cycles have environmental impacts related to material sourcing, manufacture and disposal.

#### Implement Physical and Fiscal Restrictions

The restrictive measures can be restricted access such as implementing Bus lanes, reducing convenience by preventing nearby car parking or by increasing cost for use (road charging, congestion charging etc.) Additional cost can be considered to be beneficial to those with responsibility for road provision and maintenance. Road charging can also resolve a long standing problem for urban administration groups (e.g. Town and City Councils); urban buildings generate revenue to the local administration through rental and/or taxation whereas roadways are “dead space” that cannot be used to generate income. Cost penalties against private use vehicles have a greater effect on the less wealthy as they represent a higher proportion of income or available funds. This can lead to a socio-economic class division leading to a belief of “second class citizen” and “second class transport system”.

#### Increase multi-occupant vehicles and other public transport.

Public transport systems are economically viable in high population density regions especially where there are multiple interactive services, relatively limited routes/destinations per passenger and restrictive measures against private use vehicles. Combustion powered multi-occupant vehicles generally have high fuel use; continuous low occupancy running may increase environmental emissions per passenger above private vehicles. In urban areas with smaller population or lower population density there is an increase in the routes per passenger ratio and lower average occupancy per journey. The economic sustainability of some services based on direct revenue generation is highly questionable as is the environmental impact. Section 3.3.3 and Table 3.4 show that busses consume more road capacity especially at lower speeds, when taken in combination with reduced manoeuvrability, greater road width requirement and dedicated lane allocation it can be seen that multi-occupant vehicles can responsible for congestion due to road occupancy and for extending the duration of congestion.

#### Increase road capacity through widening, bypassing and speed alteration.

This is rarely applicable to urban areas due to space availability; the number and size of buildings adjacent to the road is reduced to accommodate the increase in roadways or roadway dimensions. This may result in fewer domestic dwellings in suburban areas or fewer retail opportunities outlets in major urban centres, both of which can reduce the income for local administrations.

The need to increase road capacity emanates from the desirability for travel to a destination. The access to the destination is restricted whilst the work is completed which increases congestion. There are studies that show that once the work is completed congestion reduces but then rapidly increases again due to the desirability of the destination remaining constant while the ease of access improves; this is known as induced traffic. [100]. Bypassing may be beneficial for residential occupancy but can be economically damaging for business centres.

#### Increase road capacity through parallel routes.

Road planning involves controlling traffic flow which is only possible if traffic is confined to predetermined routes for example ring road systems around city centres that avoid through traffic. If these requirements were relaxed, traffic could utilise multiple parallel routes to the same destination. If the same volume of traffic is presented across multiple routes, the traffic density will reduce. The major concern to this is that many of the parallel routes will pass through residential areas with smaller streets. While the use of parallel paths is possible, the socio-environmental impact of using unsuitable paths makes this approach unappealing while the lack of traffic control could result in reduced traffic flow. The concept of improving the controlled use of suitable parallel paths through signposting is discussed later in this work.

#### Temporal segregation of road users

The volume of traffic is not constant over a period of time; roads are not used at or above capacity at all times of the day and night. If all of the road users were to be allocated a specific time of use, the volume of traffic could be evenly distributed over the day (or week / month) and the peak volume of traffic would reduce. The problem with this approach is that people travel at given times for a reason and that reason does not exist at other times. An example would be the starting time at a place of work; if the start time is 9:00am then travel will take place for a moderate period before this time, significantly before would not be generally acceptable and afterwards would be totally unacceptable for the employer.

Within this option there are also possibilities of staggered start times and flexible working hours. Both concepts are feasible and used but only in certain circumstances, typically where there is minimal inter-reliance on a job function or external / interactive coordination required. This would not be suitable at a school, as an example; each student and teacher must coordinate to arrive at the same time for a class. Taken in isolation this may be achievable however this may also rely on the parents of the students coordinating with the class and the employers of the parents also coordinating. If the parents are engaged in a customer facing role, the customers may also need to coordinate. This is a small example of the follow-on effects of a change to working hours to accommodate transportation change.

It is possible for some people / professions to dictate rather than accommodate working hours such as Medical Doctors or high value Mass Entertainment Industries such as Sky Sports; the latter can dictate the starting and finishing times of football matches to better suit their advertising revenue profile rather than as a benefit to society. The change to traditional football match schedules can have a significant impact on local transportation where the traffic lights have been synchronised to the anticipated traffic flow.

The concept of temporal segregation is expanded further in section 4.4 which discusses overlay traffic.

#### Refine and optimise traffic signal control to maximise traffic throughput.

Optimising traffic flow through control signals is highly desirable. It has been proven that “vehicle actuation” can be far superior in terms of traffic flow when compared to fixed time control. The major problem of fixed timing is that it requires continual tuning to account for changes in traffic requirements over relatively long periods or for change of area use or occupation. Optimal timing is different for different traffic saturation levels and, while several schedules can be established to satisfy regular occurrences such as rush hour, fixed timing plans do not adapt to unanticipated events such as accidents. Vehicle actuated systems generally fall back to fixed timing plans during traffic extremes when roads are empty, congested or have detector failures.

#### Remove all controls and rely on human nature and due consideration.

Shared-spaces or Open-spaces projects are suggested by some to be a means of allowing congestion to somehow manage itself. The underpinning theory is that if all of the “space” users act co-operatively, the overall speed may reduce but the overall flow will increase and therefore congestion, by definition, will reduce. In its most native form shared-spaces involves removing all traffic controls, road signs, kerb edges, pedestrian crossings and road markings – anything that can imply a priority of one road user over another. Removing the psychological “legal rights” to an area of road space is intended to make people more considerate of each other. There are video clips of Poynton, Cheshire available as created by an advocate [25]. Town council minutes show that the consideration was not a long term effect [24]. The reasons for the short term benefit need further consideration but it is postulated that effect is based on several factors:

* Vehicle operators driving more slowly. This may be because they are unsure of the applicable regulations or the correct protocol to adopt in the unusual or unfamiliar circumstances. It could also be a psychological view that they are somehow trespassing in a pedestrian enclave.
* Drivers are more likely to defer priority to other road users when they are travelling slowly. In Game Theory terms, the drivers are gaining a moral, social or psychological benefit that they perceive to be of greater value than the perceived time penalty incurred.
* As the environment becomes more familiar either directly or by virtue of it becoming a more common solution, drivers no longer have the view that they are trespassing and start to view other “space” users as being the intruders. Driving speeds increase. Drivers view the time lost from higher speed as being of greater value than the moral benefit of differing priority and are therefore much less likely to participate in mutual co-operation.

A related concern with regards to this method is the effect of the transition speed of the vehicles. Slow and varying rates of vehicle movement are likely to require continuous throttle and clutch adjustment, this could lead to increase environmental hydrocarbon based pollution and increased particulate pollution for the clutch plates. “Low speeds” implies the use of low gear ratios in vehicles and as a consequence less distance is being travelled for the same amount of fuel use which again increases the relative level of pollution. Finally, at low vehicle speeds there will be minimal mechanical air stirring; pollution is less likely to disperse from the built up area where there may also be minimal air movement due to sheltering.

#### Enhance technology, remove all human factors from the decision process.

This is currently the class of vehicles commonly referred to as autonomous or driverless, though there is an increasing move towards the term “highly automated vehicles” which is more suitable. Technology, economics, legal issues and social acceptance are some of the topic areas that need to be resolved. In the context of this project, if the same number of vehicles exists and try to access the same amount of road space, then congestion will still exist. Providing v2i information to a traffic control point presents the same limitation as existing control systems. The v2i will not be able to identify a clear exit as it is a contradictory paradox, either the vehicle is present and able to send a v2i message that the exit is not clear or the exit is clear and there is no v2i transmission. No indication regarding the status of an exit means that the decision regarding priority will still be made based on a backpressure system and delays will still exist. Preventing delays at a given crossroads junction would require slower vehicular movement and greater inter-vehicular gaps which will reduce the road capacity and increase congestion.

This is the most involved solution with argument and counterargument, claim and counterclaim as to the potential benefits that could be realised. While the specifics of the topic are beyond the scope of this work, the aspects that relate directly to congestion and traffic management deserve due consideration.

First consideration is the often cited “fact” that an automated system will be able to react faster than a human counterpart. The discussion in section 3.3.5 and the information shown in Table 3.5 are clear that there are at least two distinct stages that are necessary when stopping: thinking and braking. It can be argued that these can be further subdivided into observing, assessing, concluding (1), response selection, concluding (2), reacting, application and effect/outcome.

Table .14: Extended Stages of Action Determination

|  |  |
| --- | --- |
| observing | Situation awareness, registering of all potential actors. Considering historic and current states to determine movement |
| assessing | Considering current and future (predicted) state of all potential actors to determine likely actions and interactions |
| concluding(1) | Decision regarding whether an actor or interaction with an actor is likely to require a change to (our) current status / situation / vector |
| Response selection | Selecting from a range of responses the most suitable response to the conclusion drawn (e.g. no action, reduce speed, stop, alter trajectory) |
| concluding(2) | Decision whether to apply any response other than “no action” |
| reacting | Physical co-ordination / reconfiguration necessary to apply mechanical response |
| application | Application of determined mechanical action |
| effect/outcome | System response to applied action |

It is accepted that a computer / mechanical system should be able to react and apply an action faster than a human counterpart. This is mainly due to the fact that there will be hardware / actuators dedicated to a specific task e.g. a brake actuator. This hardware has a far superior reaction time due to:

a) dedicated to single function as opposed to human leg moving between the functions of acceleration and braking

b) optimised operational location within the system i.e. it can directly increase brake fluid pressure as opposed to a human operating a pedal acting as a lever to increase the force applied to a cylinder which increases the pressure in the system. The action of the lever is to increase the applied pressure on the cylinder at the expense of requiring greater travel in the pedal, greater travel taking more time to achieve.

c) Single actuation excitement, purpose built operation, pre-determined and repeatable response. The actuator is specifically designed to respond to an excitement system that is known in advance and is totally repetitive as opposed to the brake pedal system that must function over an unknown wide range excitement determined by factors including the limb length, seat position and muscle strength of the driver

The speed of the other factors are not as clearly determined. Most obviously the effect/outcome is ultimately determined by physical constraints e.g. there is a maximum rate of deceleration. As discussed in section 3.3.5 the rate of change at any given velocity is mainly determined by the friction between the brake component materials, friction between the road surface and tyre, inclination of the road and the mass of the vehicle. These factors are consistent regardless of the actuation system. Furthermore, Table 3.5, Table 3.6and Table 3.7 all show that the braking time / distance is the dominant component of the overall stopping distance.

Possibly the most controversial consideration is the ability of a computer based system to determine specifically if a response is required in a shorter period of time than a human counterpart. While is possible to compare reaction time from the precise time that a decision to react has been identified, it is not easy to compare the length of time taken to reach the decision as it is impossible to determine when the human counterpart began to assess the situation or when the assessment concluded. Finally, the criteria being assessed may not be the same. As an example, if a pedestrian were to be observed walking towards the edge of the pavement, there are many stages that could be determined to pose an actual threat e.g. moving towards the pavement edge, stepping into the road, approaching the current path of the vehicle, stepping into the path of the vehicle. Besides the physical location, there is also a temporal dimension to consider for example being in the path before, simultaneously with or after the vehicle is an obvious consideration with two definite and one open outcomes.

# Traffic Modelling

## Traffic Presence

The first observation that must be made is that the level of traffic present on a road at any given time has a purpose for being there. The most successful way to reduce traffic and congestion is to remove the purpose for the traffic however this is unlikely to be a preferred option in the majority of cases. Section 3.4.3 has already alluded to the fact that traffic presence and queueing can be an indication of economic vibrancy for an area, removing either the traffic or reason for the traffic is likely to have an opposite effect, indicating local economic downturn.

### Traffic Efficiency

A common reason for traffic to exist is as a means of transportation, a method of moving “something”, normally people or goods. The efficiency of transportation can be measured in many ways and is dependent on the overall objective. For example, for emergency services the efficiency is normally represented primarily in terms of time whereas for goods it could relate to number of items, weight/volume of product and for public transport it would be the number of passengers. Regardless of the primary factor, other factors include the distance travelled, time taken, amount of fuel used and cost in terms of fuel, vehicle maintenance and labour. It is possible to continually extend this list to create entire research projects [101], certainly lobbying bodies [102] and large transportation companies invest significant time and resource into minimising environmental damage and overall operating costs by maximising efficiency.

To create a foundation for this section, a model is developed that considers the effectiveness of a transportation system. Beginning with the simplest considerations, the number of “things” that are transported a distance for a given amount of fuel.

Equ( ‑)

Observations from Equ( 4‑1) can be made regarding efficiency for example efficiency, η, can be improved by using less fuel, moving more items or moving a greater distance. The efficiency will improve while only one variable is changed in the way described or while Equ( 4‑2) remains true.

Equ( ‑)

The efficiency will tend to ∞ if fuel used tends to zero, this is a very important concept if the “cost of fuel” is considered rather than the actual unit of fuel. This is one of the driving forces towards the use of renewable, self-derived energy sources such as private wind or solar farms. Within the context of transportation *cost* efficiency, energy produced at zero cost means that the efficiency remains high even when only a few items are transported. This assumes that the distance between the source and destination does not change. This in turn allows a larger number of smaller vehicles to be used which maximises the load distribution capacity when the load sizes vary over time or consignment. There would be a highly beneficial increase in the distribution capability for the haulage company with negligible vehicle cost increase. The negative impact of this approach is an increase in labour costs and an increase in number of vehicles, traffic and congestion. If fully autonomous vehicles were to become a reality, this could become a viable option for transportation companies as it would reduce the labour cost for a highly dynamic fleet. This could be further evidence to support the view that driverless vehicles will increase rather than decrease congestion.

Equ( 4‑1) also shows that efficiency tends to zero either when nothing is transported or when no distance is moved. This is a simple mathematical proof that a stationary vehicle is totally inefficient in terms of transportation. The efficiency equation only relates to the items that are being purposely carried between two points, i.e. we can discount any item that relates to the vehicle. Furthermore, not all of the distance travelled is beneficial i.e. the effective distance moved is a direct line whereas the actual distance includes only accessible paths (roads, rail routes, ferries and flightpaths). On a larger scale this can also consider curvature of the Earth, which could be avoided by the use of “Gravity Trains” or direct point to point subterranean links described by various authors [103]. The direct-line concept is outside of scope and though mentioned for completeness, it is unachievable due to practical physics and geology. For example, P W Cooper discusses a link from Boston to Washington D.C. and suggests a depth of only 5 miles however the single deepest borehole depth that has ever been constructed is 7.5 miles [104], one of the limiting factors was the well temperature of 180°C / 453K. For context, the continental crust of The Earth is between 18 and 28 miles [105] and the average diameter is approximately 7900 miles.

The original equation, Equ( 4‑1), can be generalised to energy used over a period of time and include only items purposely carried over the necessary distance:

Equ( ‑)

Combining the previous paragraph with the content of Equ( 4‑3), it should be clear that a taxi or bus without a fare has zero transportation efficiency regardless of the occupancy of a driver. This is an important concept when considering one of the proposed benefit of driverless vehicles in reducing traffic congestion. A further observation is that a direct route is more efficient. Route efficiency is not under discussion but, again for completeness, it can be observed that

Equ( ‑)

Finally, Equ( 4‑3) does not include a reference to the type of fuel used, only the power that is required to move a given mass through a given distance / elevation. Different fuels release different amounts of energy for the same given weight or volume. Inefficiency or system waste occurs when there is an unnecessary transition of energy from one form to another including the undesirable by-products of transition. When considering the car purely as a transportation system, inefficiency considers the contrast between the useful work obtained through conversion of stored energy into useful motive power against the technically unnecessary outputs including heat, light, noise and the undesirable by products of chemical conversion. For fossil based fuels the arguments are well known however similar arguments exist for most, if not all, currently available portable energy sources. For example, secondary storage cells (rechargeable batteries) have a finite life [106] and generally contain toxic components which are hazardous to the environment and expensive to recycle [107], a 2012 report conducted for the California Air Resources Board concluded that vehicle battery recycling is not economically feasible [108]. The report also shows that while CO2 is much lower for battery vehicles, it is still a very significant amount. Life Cycle Analysis (LCA) of solar photo-voltaic cells undertaken by Sherwani, Usmani and Varun gave CO2 figures for solar produced electricity ranging from approximately 10 to nearly 300 g/kWh [109]. While the energy source of sunlight may be considered as renewable, the methods of collection, storage and use are clearly not. This totally justifies the argument presented by Equ( 4‑3), i.e., within the context of this project, the type of fuel is irrelevant. A vehicle which has motion restricted due to congestion is fuel inefficient; in this situation any energy used (including heating, cooling, lighting, entertainment) can be considered as wasted and is needlessly contributing to CO2 emissions regardless of exact timing and location of fuel derivation. This is important when considering motive efficiency such as the ratio of fuel required compared to the amount of load carried but by considering only the energy used, we can dispense with any immediate discussion regarding the motive power derivation that is we can preclude any debate over the most suitable motive power source and do not need to compare diesel, petrol, LPG, fuel cell, battery and so on.

Equ( 4‑3) is not intended to provide a solution to a problem, it is intended only to identify the concept that there are different levels of benefit that can be derived from transport methods and to provide a simple method of comparison. The equation does not fully encompass the entire dynamics of vehicle efficiency, only that there is a range that extends from some arbitrary maximum point and tends to zero when no useful work is being completed or the time to complete becomes infeasibly large, for example due to congestion.

## Traffic Flow

Measuring and controlling traffic flow is a very complex problem. Consideration must be given to type of road, type of traffic, method of control, tidal flow impacts, temporal effects, convergence, divergence, sources, destinations and cross flows.

It is commonly observable that traffic flow is tidal in nature, especially in urban and inter-urban roadways. The explanation for this is that the population of developed nations (those with the highest number of vehicles per capita), generally live in residential areas away from the centres of mass employment. There are potentially a large number of socio-economic reasons for this including:

* the undesirable negative impact of pollution and particulate fallout from industrial production meaning that few people prefer to live near the industrial area
* the preference of employers to be able to draw from a larger population base ensuring a better choice of suitable and economic workforce
* the need for commercial centres to be able to attract a larger and more diverse customer base than the immediate locality

It should be noted that this simple list identifies that the choice and preference of segregating commercial and domestic regions is not unilaterally beneficial to the detriment of other parties. The absolute reasons for urban sprawl and diversity in use of land are more suited to sociology and generally outside of the scope of this work other than to recognise the impact that this has on traffic flow modelling.

The next point is the concept that most businesses have fixed starting and finishing times. It can and is argued that staggering of times would reduce traffic congestion however the consequences of this approach are not trivial to resolve. Large numbers of the population are engaged in tasks that have sequential activity that require prior knowledge of the subsequent task e.g. taking children to school while travelling to work. Many commercial businesses rely on the availability of supporting or competing businesses as a means to draw customers to an area e.g. shoppers will go to a shopping centre when the know that all of the outlets that they want to visit are open; staggering opening and closing times would be highly detrimental to sales and would result in a “core hours” opportunity which would directly lead back to fixed opening and closing times.

Accepting that regions are primarily utilised for different purposes, we can now consider the effect that this has on traffic flow. Figure 4.1 provides the general idea, traffic flows from “home” to “work” in the morning rush, remains there for the duration of the working period and then flows in the opposite direction in the evening.

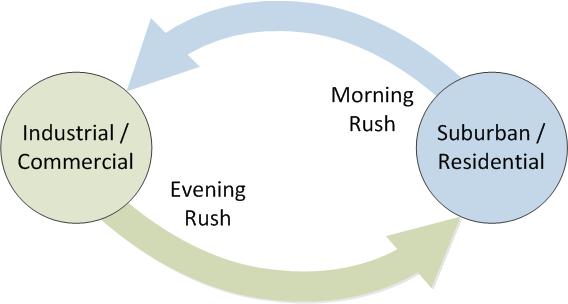


Figure .1: Concept of Tidal Flow

The incredibly simple diagram shown in Figure 4.1 can be used as the basis for many concepts, the first of which is considering traffic light timing. Section 3.4 discussed traffic light timing and showed how traffic lights cause breaks in traffic flow that leads to delays but only considered the minimal sequencing given in para 2.3.2. However, many more sequences are possible; one suggestion is given by Figure 4.2 and Table 4.1. It must be clear that each sequence stage will require the full transition through the stages of Red, Red/Yellow, Green, Yellow. The “Input” columns and “Totals” row in Table 4.1 are intended to illustrate how traffic lights can be used to allow for Tidal flow. In the original table each direction has the same number of allowed inputs (6) across the entire sequence of 24 inputs, we can conclude from this that there is an equal period for each stage i.e. 25% of the cycle time.

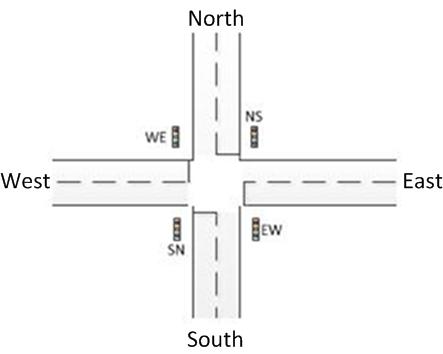


Figure .2: Simple Crossroad Junction

Table .1: Possible Junction Sequences

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sequence | Flow allowed | Input from N | Input from E | Input from S | Input from W |
| 1 | N→S, N→E, S→N, S→W | 2 |  | 2 |  |
| 2 | N→E, N→S, N→W, W→N | 3 |  |  | 1 |
| 3 | E→W, E→S, W→E, W→N |  | 2 |  | 2 |
| 4 | E→N, E→S, E→W, N→E | 1 | 3 |  |  |
| 5 | S→W, S→N, S→E, E→S |  | 1 | 3 |  |
| 6 | W→N, W→E, W→S, S→W |  |  | 1 | 3 |
| Total |  | 6 | 6 | 6 | 6 |

Now if we assume that the residential area of Figure 4.1 is to the East of the junction and the industrial area is to the West, we can adjust the morning, normal and evening timing to give higher priority to the direction with the greatest demand. For example we could allocate two additional time periods to sequence 4 in the morning (and similarly to sequence 6 in the evening). The result is shown in Table 4.2; neglecting the all red transition period traffic approaching from the East now has 40% (12/30ths) of the junction time while each other approach has reduced to 20% (6/30ths). If this was a typical 90 second cycle, the East approach would have increased from 22.5 seconds to 36 seconds while all other approaches would reduce to 18 seconds.

Table .2: Unequal Time Allocation to Improve Tidal Flow

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sequence | Input from N | Input from E | Input from S | Input from W |
| 1 | 2 |  | 2 |  |
| 2 | 3 |  |  | 1 |
| 3 |  | 2 |  | 2 |
| 4 | 1 | 9 |  |  |
| 5 |  | 1 | 3 |  |
| 6 |  |  | 1 | 3 |
| Total | 6 | 12 | 6 | 6 |

This example only considers a single isolated junction however it would be more realistic to consider a sequence of unequally spaced junctions would exist between the residential (Origin) area and the industrial (Destination) area. Section 2.3.4.3 introduced coordinated traffic signals and subsequently identified an issue in so much as to state that traffic approaching the co-ordinated area from one direction will benefit from the coordination while in the opposite direction the flow is effectively disrupted and penalised. The application of coordinated traffic controllers is based around the need for tidal flow, with the bulk flow predominantly in one direction at any given time. There have been many papers written on controlling the flow of traffic under tidal conditions which are commonly described as “platooning” or “green-waves” [110] [111] [112] [113]. Some papers refer to the possibility of bi-directional green-waves with a claim that the use of vehicle communication systems and centralised control can enable this [114] however, as shown in 2.3.4.3, this is not thought to be realistic unless there are very specific conditions including under-saturated flow, central division (dual carriageway) and absolute certainty regarding the source and destination of the wave. In the paper [114] Tomescu presents some interesting points but does not consider the wide variety of human behaviours that can lead to green-wave breakdown or the flow of traffic from the minor directions. The breakdown of green-wave flows has been studied in depth by authors such as Kerner [115] [116] and Castillo et al. [117].

Castillo et al. [117] discuss resonance within green-wave and consider the implications of the experiment initial conditions of pre-existing traffic jam. They assert that the controlling factors change as the system approaches resonance. They further assert that a simple cellular automata model is sufficient near the resonance condition as the variable effects of acceleration and braking do not appear to have significant influence. The assumption that can be drawn from this idea is that the authors are identifying that when the flow approaches resonance, the vehicles can be considered to move as a single homogenous block rather than a cluster of independent bodies. This is one of the defining principles of green-wave movement. The paper does observe that driver responses are not deterministic but does not consider flow in the opposite or conflicting direction. Ultimately the paper discusses the capabilities and uses of cellular automata as a means of traffic modelling.

Clearly defined and cyclically repetitive tidal flows are the basis for most forms of predictive traffic management methods. As described in section 2.3.5, adaptive traffic signals such as timed or MOVA require frequent survey and update to ensure that they remain current; SCOOT systems should adapt over time but will not respond to rapidly changing circumstances or isolated incidents.

The flow rate of a motorway under normal conditions should vary relatively slowly over time with peaks building and receding over a daily / weekly cycle although the rate may become rapidly time variable beyond the critical density point. In an urban road network, controls and interactions with other road users mean that even outside of a daily / weekly cycle the flow rate can be rapidly time variable but can become slowly time variable beyond the critical volume.

The primary interest for this project is the urban road network as this regularly affects the largest number of users; in an urban area there are more road users making shorter journeys at lower speeds. Urban networks have the highest number of controls applied and therefore the largest scope for change; the impact of any change will be more significant and more widely applicable.

## Traffic Types

The technical challenges of traffic control and congestion management are exacerbated by the variety of transport methods, legal and ethical control and the activity and operations undertaken. For example, the transport method could be a motorised or non-motorised mechanical, pack animal or some form of pedestrian system. Table 4.3 shows a list of common methods and typical uses, this is not intended to be an exhaustive list but an indication of variation.

Legal and ethical control can include operator restrictions, fitness for purpose, safe operation and interaction with other transport modes. This is intended to highlight the differences between, for example, Heavy Goods Vehicles (HGV) and completely unrestricted animal life, the former having operator age restrictions, increased training, increased assessment testing, enhanced vehicle maintenance checks etc. while the latter is completely autonomous in activity and behaviour with no moral or legally enforceable control. It is not uncommon for most, if not all, of the major groups shown to be present in the same road space at any given time.

Not mentioned in the table are other items that may be in the roadway that may impact road users such as road surface damage or litter and other inanimate items propelled by unrelated forces e.g. wind, gravity or pedestrians not directly using the road (children playing nearby)

Each transport type makes both cyclic and non-cyclic journeys. Examples of cyclic journeys include travelling to a place of work, completing a delivery round or a timetabled bus route. Examples of non-cyclic events could be visit to shops or friends. Cyclic events are generally considered timetabled however an appointment with a Doctor could be both timetabled and non-cyclic. Random variation of a few seconds in the starting time of even strong cyclic events and/or the random action of other controlled or uncontrollable road users mean that a journey is unlikely to be identically repeated. This is covered in further depth in section 2.4.1 (Chaos Theory) and section 4.5, Why the Bus (or Train) is Always Late

Table .3: Variety of Common Transport Methods Observable on Urban Roads

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Major Group | Subtype | Attributes | Negatives | Legal (ethical) requirements |
| Motorised Mechanical (1) | Car, Light Van | Fast, wide speed range, adaptable, long distance capability. Multiple occupant | Overly comfortable?, operator easily distracted, significant numbers. | Use, maintenance |
| Motorised Mechanical (2) | Motor-cycle | Fast, wide speed range, agile, moderate distance. Limited occupancy | Fast, relatively small, minimal protection. | Use, maintenance |
| Motorised Mechanical (3) | Goods Vehicle (HGV, LGV) | High load carrying capability, moderate speed, moderate (restricted) speed range, long distance. Low occupancy | Poor local visibility, difficult to optimise stability (and braking) for all loads, long acceleration and deceleration distances | Increased requirement on use and maintenance especially hgv |
| Motorised Mechanical (4) | Passenger Vehicle (PSV, minibus) | Moderate speed, moderate (restricted) speed range, long distance capability. High occupancy | Frequent stops, driver monotony and anxiety, long acceleration and deceleration distances | Increased requirement on use and maintenance especially psv |
| Non-motorised mechanical | Bicycle (functional, fitness, racing), skateboard etc. | Low cost, low speed, relatively wide speed range, short distance. Single occupancy | Relatively small, vulnerable, minimal protection | Minimal use or maintenance |
| Pedestrian | Walk, run, wheelchair. With or without load | Free, very low speed, short distance. | Small, vulnerable, minimal protection | Minimal use |
| Biological Load bearing | Horse, Mule | Low speed, limited speed range, short distance – typically rural use. | Vulnerable, minimal protection, control may be affected by other vehicles | Minimal use |
| Biological non-load bearing | Dog, Cat, Indigenous wildlife | Low speed, relatively wide speed range, typically urban | Vulnerable, no protection, actions may be affected by other vehicles, may disregard all other vehicles | Negligible / none |

## Converging and Diverging Traffic

Section 1.8 introduced some different road types in very broad terms and traffic flow was discussed in section 4.2. The notion of tidal traffic was considered in respect of origins and destinations. The exact nature of the source and destination for traffic might however still be unclear.

First consider a focal point destination such as a place of mass employment. Using the basic concepts of Braess Paradox [46], a popular destination necessitates good access. Braess goes on to say that increasing access makes the destination more attractive and attracts more traffic which increases popularity thereby increasing traffic and congestion, however if we assume that there is a fixed size of workforce then we need only consider the initial stage. The mass-employment concept means that the traffic will converge at a single temporal as well as physical point. Taking the entrance to the mass employment facility as also representing the exit from the approach road we can understand that this will be the highest concentration of traffic for this destination. As a mass employment centre it is probable, if not certain, that the workforce will approach from several different residential areas and therefore at some distance away from the destination there will be one or more junctions where the traffic combines. The concept is shown in Figure 4.3 where traffic presented to destination D is supplied by route A. The route A is conservative meaning that there are no losses or gains along the route, the traffic volume at one end is the same at the other end but may be displaced in time. Junction J1 combines the traffic from each of the routes identified as B and because the traffic is conservative the exit of J1 has the same traffic volume as Route A. Similarly, the junctions J2 combine the traffic from the routes C. This illustration implies equal traffic from all directions however this does not need to be the case, the traffic on route A equals the sum of routes C at a slightly displaced period of time. For the purpose of this illustration, the lengths of the routes are not important even though they would be highly significant in terms of providing buffer space in a real world context. The Junctions in the diagram may be traffic island, traffic light or stop/give way, the concept of combining is the most important factor.

The other concept for converging traffic occurs at the origin, in this case the residential part of the journey. In a typical residential area in the UK there will be a number of different types of roadway that interconnect and then link to a minor arterial route and eventually a major arterial route. Designing the residential road network with physical access restrictions such as “traffic calming” reduces the likelihood of through traffic whilst having only a single entry/exit ensures that the traffic will be light residential. A stylised view of the residential concept is shown in Figure 4.4

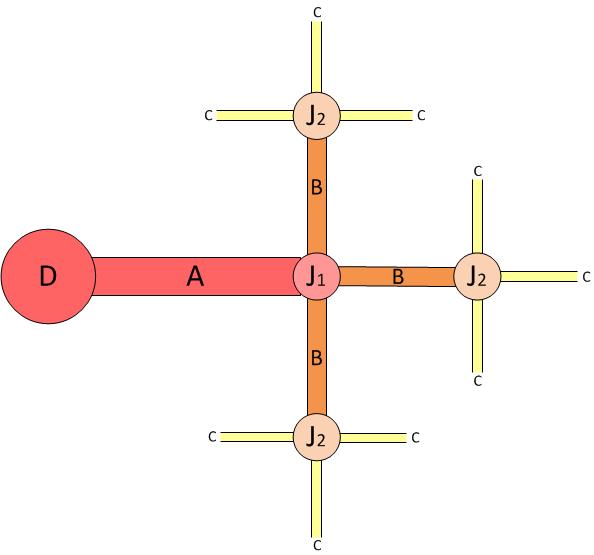


Figure .3: Traffic Converging at a Destination

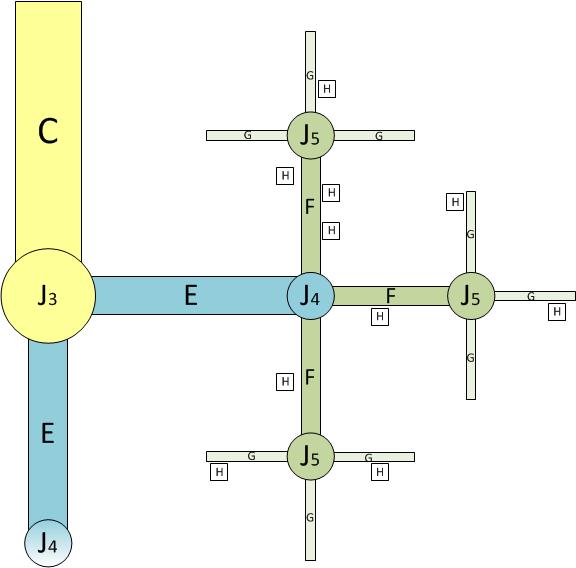


Figure .4: Traffic Converging from an Origin

The route C in Figure 4.4 is intended to represent the arterial route C in Figure 4.3 albeit forming a very compressed city structure. J3 combines the traffic for the residential access routes E, J4 combines traffic from the major residential streets F and J5 combines the minor residential streets (typically cul-de-sac or with only a single vehicle entry/exit). The small “H” boxes indicate residences of interest for the given route, i.e. places where a vehicle will enter (or leave) the traffic flow intending to arrive at a given destination D at a given time. Not every resident of any street will want or need to make the journey to the given D at the same time or even at all and therefore the number of H on any street can be zero even when there are numerous residents and vehicles. It is possible for there to be more or less H boxes in any area or on any street type and they may exist on any of the roadway types A-G.

To clarify the convergence concept, consider D to be a mass employer with a work start time of 8:00am. The employees must leave their individual origin points, H, to arrive at D at e.g. 7:50am. In reality this will be a range of times with a normal distribution curve but we only need to consider the peak arrival time. Vehicles will leave each point H at a given time that they have determined by experience to provide the best opportunity of arriving at D at 7:50am on any given day. This experience will take into account distance to travel, anticipated traffic level and expected traffic flow. Vehicles will leave H at individual times but all vehicles within a small geographic area, e.g. those all connected to a single junction J5, will begin the journey at very similar times.

At any point in time, each junction J5 will combine the number of local origin H intending to arrive simultaneously at destination D onto each roadway F. The volume of traffic on roadway in any period will be the sum of the traffic at each local G in a previous period plus any traffic originating on F. At some slightly delayed period, each junction J4 will combine the local roadways F onto roadway E. It is important to observe that because the given junction is a fixed physical distance and, at any time, an approximately consistent temporal distance from destination D, all traffic within the flow will arrive at the given junction at a similar time. These are very important concepts when considering traffic modelling with Cellular Automata or Macro simulation. The process continues until the traffic eventually converges at roadway A, travels to Destination D and is then, in terms of traffic flow, ceases to exist. The convergence at A occurs at a given time (in this case 7:50am) regardless of the distance travelled or the journey starting time at H.

This results in an overlay network as shown in Figure 4.5. This version of the overlay considers three residential areas and a destination D which is a single site of mass employment in an isolated industrial area. The overlay can be placed directly over a map to show a general movement or can be fitted to the primary roadways to show expected flows

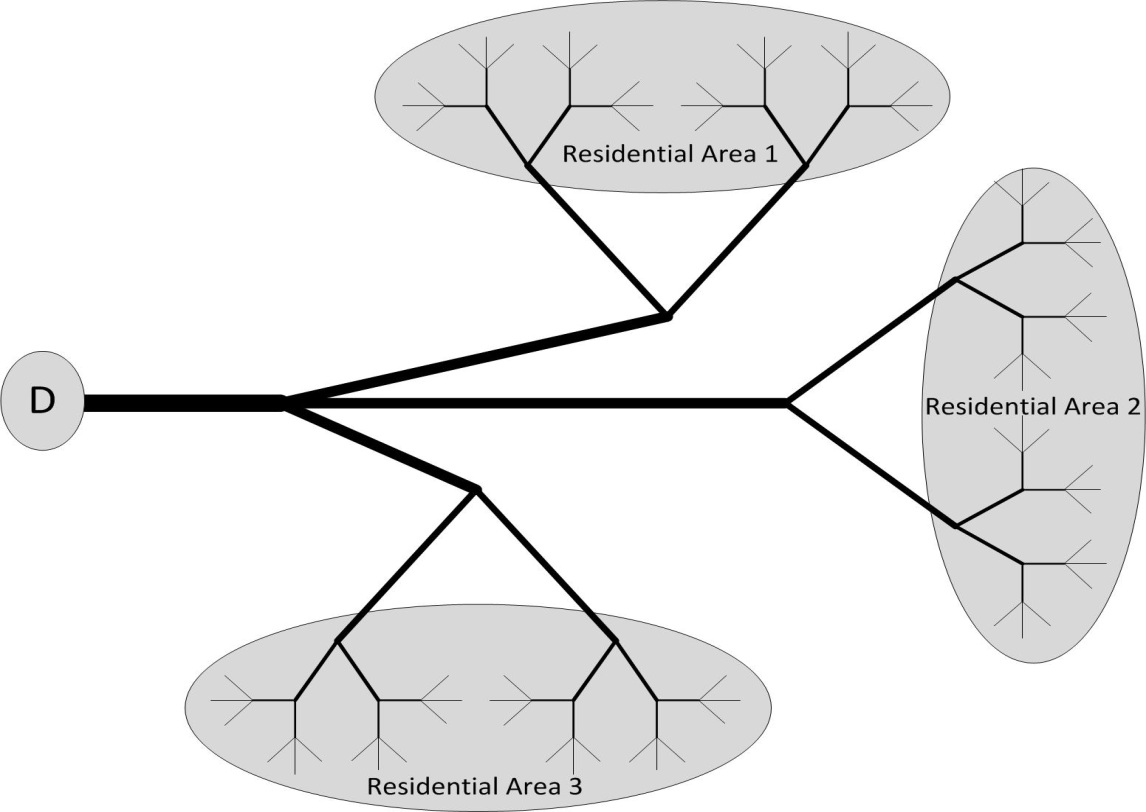


Figure .5: Overlay Concept

The traffic shown by the overlay is fixed however the flow of traffic must be shown in a 3-dimensional overlay with traffic volume plotted against each road over the time period of interest. An example of a stacked overlay of the “Residential Area 1” roadways is shown in Figure 4.6. Time period 0 shows the minimal traffic leaving on the “G” roadways, at time period 1 the traffic has moved on to the “F” roads, then the “E” and finally “C”. The figure clearly identifies how the apparently negligible amount of traffic originating in the residential area rapidly builds to a high density over a short period of time due to funnelling or convergence. At time period 5, the traffic intended for destination D has cleared the area.

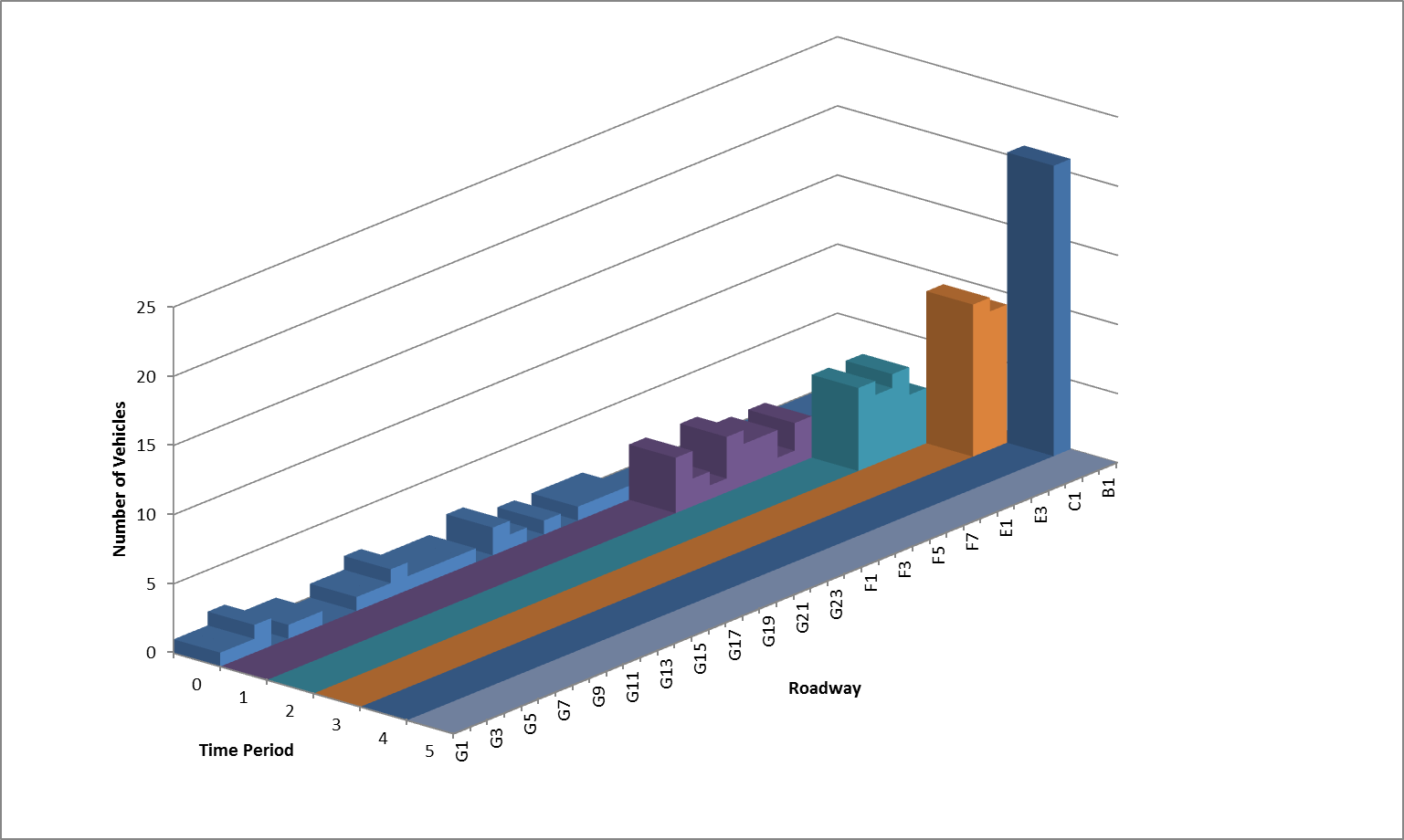


Figure .6: Time Stacked Overlays

The time stacked overlays in Figure 4.6 are an example of macro modelling of traffic flow. It shows very clearly that the traffic moves as an apparent block from one area to the next even though the blocks consist of individual vehicles. This simple model implies that all traffic will arrive and be serviced simultaneously at each junction and on each roadway. In practice this is not possible as each vehicle must be serviced sequentially at the junction and along the roadway. The maximum number of vehicles concurrently serviced is determined by the number of complete independent parallel paths available i.e. the number of lanes, car park entrances and so on. Regardless of practicality, traffic flow can be implied using this method of macro modelling.

Using macro traffic modelling, the physical overlay from Figure 4.5 can be reduced in complexity. Each roadway can be considered to be an individual system with an origin and a destination and simple arithmetic is used to determine the loading on the subsequent road section. “Residential Area n” becomes a source number while the product of road length and speed become a time delay before the source number moves to the next section. In applying this method it becomes clear that the network can have many sequential destination points. The destination point can also be considered as a “virtual destination” which is the point where the traffic leaves our management area or traffic domain. A virtual destination could be a major inter-urban trunk road into another region that can sink all of the converged traffic offered. This concept was presented earlier in Equ( 2‑3), where it was stated that there will be clear access and no (or reducing) queueing if the road capacity, Vc, is greater than the volume of traffic approaching, Va.

The converse of the virtual destination is the virtual source; large volume traffic arriving from another traffic domain. Gating used at this point could reduce the volume of traffic entering the local domain which should reduce congestion within the domain however this does not remove the problem, it simply moves the problem to an upstream roadway. Politically it would not be acceptable for a local domain to delay traffic a major trunk road passing through the domain without justifiable reason, economically it would be unacceptable to prevent customers and traders from entering the domain to generate commercial viability and gating can generate traffic on less preferred local routes where they offer a parallel entry point or gate by-pass opportunity (the so called “rat-runs”).

Referring back to Figure 4.4, the relatively small number of positions marked H is intended to imply that there is more than one temporal or physical destination from the general origin defined as “Residential Area n”. The example given considers an industrial setting where 8:00am may be a common start time, however the same physical destination may have other starting time for “shift workers” or “office staff”. A separate physical and temporal destination may also exist for schools, office and shop workers. The overlay for “school run” traffic will be significantly different because the temporary destination may reside inside the “residential area” and the journey is circular; the ultimate destination is the origin but shifted in time. The concept of converging and diverging flow can be seen in the school run example shown in Figure 4.7. The simplified model assumes that there is a significant period of wait at the destination that allows the influx traffic to clear in advance of the outflux traffic. In a real situation it is highly likely that the traffic will not all observe the delay and there will be an interaction between the two flows that will increase the overall journey time.

The school run overlay is similar to the city centre overlay, the destination is inside the management area or traffic domain however with the city centre concept, traffic ceases to exist inside the model in the influx period of the tidal flow and suddenly reappears at the outflux period. Within the city overlay there may be a small number of high volume destinations such as carparks but there will also be a large number of limited volume destinations such as private or on-street parking.

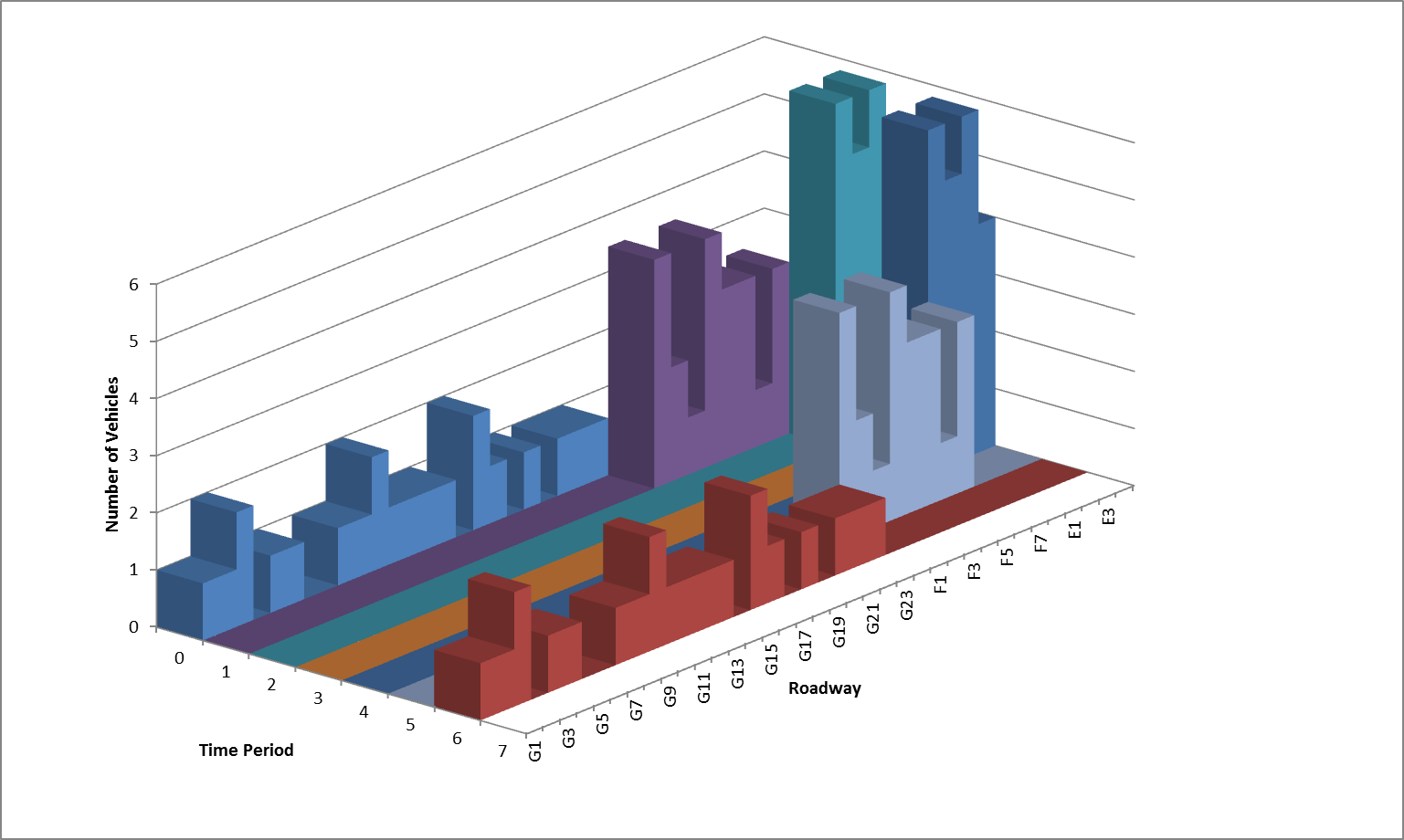


Figure .7: "School run" Converging And Diverging Flow

An example of a city overlay is shown in Figure 4.8 to Figure 4.11. The first 3 figures show individual Origin-Destination (OD) journeys while the final figure shows the combination of many such journeys. The individual overlays are shown to clarify the multiple journeys that combine in Figure 4.11, the combination is necessary because the journeys all occur in the same temporal instance i.e. the journeys are concurrent.

It is possible that any or all of the overlay situations depicted may temporaly overlay each other or may occur at different times and not impact each other. Separating the overlays would spread the traffic loading over a longer period of time which would reduce peak flows and consequent peak congestion

The overall effect of convergent traffic is that the traffic density is likely to increase towards the proximity of the destination. The converse should also be true, that the traffic should become divergent nearer to home on the return journey; traffic density and associatted delays should reduce as a consequence.

|  |  |
| --- | --- |
| Figure .:City overlay 1 - single O, D | Figure .: City overlay 2 - Multi O, Single D |
| Figure .: City Overlay 3 - Multi O, Single large D | Figure .: City Overlay - combined |

## Why the Bus (or Train) is Always Late

The problem of managing a strict timetable is apparent on bus and train routes. The largest problem is that the service is always late or at least it is apparently always late.

A part of the problem can be attributed to there being a tendency to remember the unusual events rather than the mundane ones. While not related directly to transportation, this idea is supported by the strategies of teaching and learning whereby students remember and learn more from teaching that is “different” and therefore interesting. Blumenfeld et al., amongst others, discusses this subject [118]. In terms of waiting for a form of transport, it has to be assumed that the service is actually on schedule for most of the time and therefore the exceptions (late arrivals) are more memorable. If however the service is always late, then the schedule needs to be amended as the physical constants are fixed.

The next part of the problem is that the delay has cause and is cumulative. If a delay occurs due to other traffic, the cause is very likely to be based on convergent traffic and will continue to restrict traffic flow (cause further delay) for consecutive stages of the journey. In considering bus transportation the route is the path taken to begin and return to a single point, the journey is an arbitrary number of stages along the route and a stage is the movement from one stop to the next; a passenger makes a journey along a route that consists of one or more stages. If the bus service is defined to complete at stage at, for example, an average of 20mph but the maximum speed of the other traffic is 20mph or less, a delay will occur. This is because the service is actually defined to complete a stage in a given period of time while restricted to a maximum speed. The period includes the stop time for passengers to safely leave and enter, time to re-enter the traffic flow, time to accelerate to the flow speed and to decelerate to stop. The acceleration and deceleration rates must be limited to levels that are not uncomfortable to the passengers. Further delays can be added by traffic entering, leaving or crossing the path in front of the bus.

Finally, the delays can be amplified by traffic control systems; traffic lights can only increase the delay and not reduce it. If the traffic light is showing green when the bus approaches then no delay is added but the delay is not reduced. If the light is red on approach, a further delay is added. This is the basic concept discussed in section 3.4.

There is no stage where a “negative delay” can be added. A negative delay implies that the journey can be made in less time than originally planned; this is only possible if the original plan overestimated the journey time, if the journey constraints (speed limit, acceleration and/or deceleration rate) are exceeded or a buffer time period is added at each stop.

Overestimation: If the estimated speed of the traffic on the route is lower than the actual speed then the transient part of the journey can take less time than estimated. However, the starting time for each stage of the route is set by the timetable, the bus cannot begin a stage in advance of the timetable as this is the contractual obligation of the service. Passengers who arrive on time at a stage point have a reasonable expectation of service that cannot be provided if the service has left the stage early. While it is feasible to recover lost time, this is on a stage by stage basis; the bus can only ever leave a stage on time or delayed, it can never leave in advance in anticipation of a delay that may occur at a later stage.

Buffer period: There must be an allowance made for passengers to leave and enter at each stage point. If the bus arrives ahead of the departure time then the allowance time must extend to prevent early departure. Conversely if the allowance time is longer than necessary, it can be reduced when necessary to recover lost time in any stage.

Exceeding Constraints: Where there is other traffic on the journey, it may not be possible to exceed the constraints and where there is no other traffic it might be illegal to exceed the constraints. From the general profile in Figure 4.12, t0 is the starting point of the stage, the bus accelerates uniformly to t1, maintains a constant velocity to t2 and then uniformly decelerates to the end of the stage at t3.

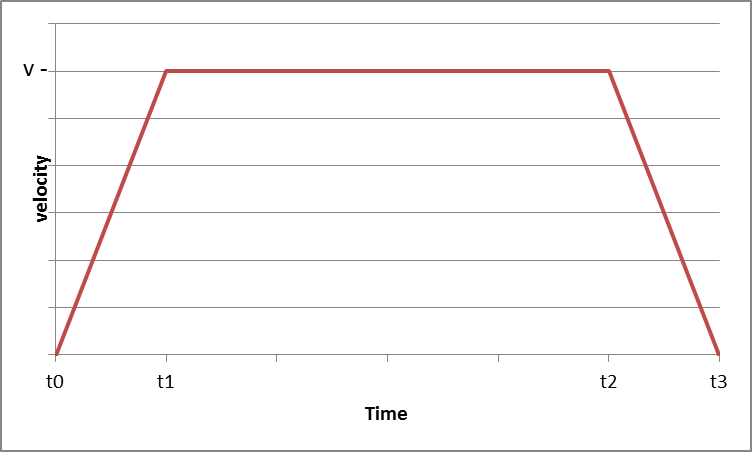


Figure .12: Velocity/Time Curve, Uniform Acceleration And Deceleration

For a given distance, stotal, a maximum velocity, vmax, and an acceleration (deceleration) rate, a, and that:

Equ( ‑)

Then the time to accelerate (decelerate) to the maximum speed is

Equ( ‑)

The distance covered during the acceleration (deceleration) phase can be calculated from

Equ( ‑)

Using Equ( 4‑5):

Equ( ‑)

Equ( ‑)

In this case, the intial condition is zero, c=0. Using the assumption that the acceleration and deceleration rates are equal, the distance covered during each phase is equal and so the distance covered at static maximum velocity is given by

Equ( ‑)

And the time taken to cover this distance is

Equ( ‑)

Finally, the average speed can be calculated from the total distance covered and total time taken:

Equ( ‑)

Using a value of a= 2m/s2 and a total distance of 1 mile (1609m), the values obtained from Equ( 4‑5) - Equ( 4‑12) are shown in Table 4.4 for a stage completed at a maximum velocity, v, at the speed limit of 30mph and in the same zone with traffic limiting the maximum speed to 25mph.

In the following tables:

v(mph) = maximum cruise speed within the stage in miles per hour (v in Figure 4.12)

v(m/s) = maximum cruise speed within the stage in metres per second (v in Figure 4.12)

taccel = time taken to accelerate from zero to v(mph) (t1 in Figure 4.12); equivalent to time taken to decelerate to zero from v(mph) (t3-t2 in Figure 4.12)

saccel = distance covered during taccel

svmax = distance covered at v(mph)

tvmax = time taken to cover svmax at v(mph) (t2 – t1 in Figure 4.12)

ttotal = total time taken to cover stage distance (1 mile = 1609 metres)

vave(m/s) = average velocity to cover stage distance based on ttotal in metres per second

vave(mph) = average velocity to cover stage distance based on ttotal in miles per hour

Table .4: Average And Max Velocity Per Stage

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| v(mph) | v(m/s) | taccel | saccel | svmax | tvmax | ttotal | vave (m/s) | vave(mph) |
| 30 | 13.41 | 6.71 | 44.97 | 1519.41 | 113.29 | 126.7 | 12.70 | 28.4 |
| 25 | 11.18 | 5.59 | 31.23 | 1546.89 | 138.41 | 149.6 | 10.76 | 24.1 |

If the time deficit incurred during the restricted flow is to be recovered on the subsequent (equal length) stage, the two stages will need to be completed in 253.4s (2 x ttotal at 30mph), therefore stage 2 will need to complete in the remainder of 253.4 – 149.6 (the time taken to complete the restricted phase) or 103.8 seconds. This means that the average speed for the second stage must be 34.7 mph without allowing any inter-stage stop time. Any time allowed for a stop period is deducted from the total travel time of the second stage.

Applying the calculations derived previously it is shown that the maximum speed in the second stage must be 37.7mph without the inter-stage stop time and 40.05mph including a 5 second stop as shown in Table 4.5.

Table .5: Ave And Max Velocity To Recover Previous Stage Delay

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| v(mph) | v(m/s) | taccel | saccel | svmax | tvmax | ttotal | vave (m/s) | vave(mph) |
| 37.7 | 16.87 | 8.44 | 71.16 | 1467.02 | 86.95 | 103.83 | 15.50 | 34.67 |
| 40.05 | 17.09 | 8.95 | 80.14 | 1449.07 | 80.94 | 98.84 | 16.28 | 36.42 |

It should be clear from the preceding calculations and explanation that it is not possible to recover lost route time without exceeding legally enforceable speed limits. It may not be physically possible to exceed speed limits if the initial delay was the result of slow moving traffic.

This work raises a paradox that public transport services will be delayed and the schedules will be unreliable during periods of congestion caused by high traffic volumes, which are the periods when there are high levels of passenger traffic. The paradox is that the delays will be greatest when the demand is highest; to satisfy the highest demand would require a more frequent service that would increase traffic levels and make the service less reliable and therefore less able to satisfy the demand.

# Detail of Conceptual Solution

## Outline

The earlier sections of this work have explained the nature of the problem and the currently applied solutions. In this section a theoretical alternative is proposed, described and evaluated. The section goes on to describe and justify a method of testing that will provide comparative results to enable the evaluation of this proposed approach.

Throughput of traffic, or flow, is determined by multiple factors including type and size of vehicle, roadway dimensions and applied traffic control method. The objective of a control method is to ensure safe interaction between vehicles with contrasting intentions at a contended space; or, in simpler terms, to prevent incidents and accidents at junctions. Traffic control methods can be broadly grouped into three different methods namely courtesy, passive control and active control. Active control in urban environments is normally achieved through the use of legally obligated indication systems know as traffic lights. The objective of a traffic light is to provide temporal segregation between transportation systems that utilise a common physical space, a form of time division multiplexing (TDM). The transportation methods can be any systems that require use of the same physical space e.g. rail vehicles, trams, busses, pedestrians, cyclists and so on. This work focusses on urban traffic controlled by traffic lights including road vehicles and, to a degree, pedestrians.

Earlier sections of this work have described how traffic flow can be influenced by the action or inaction of drivers; this solution does not change this but will offer the opportunity for the greatest overall throughput of contending traffic flows.

The main problem associated with control systems employing traffic lights is that they periodically prevent traffic flow in each direction and thereby can act to reduce overall throughput. As the number of vehicles in use has increased over many years, it has become more important that the time allocation to each direction is optimised to provide a level of “fairness” in respect of the different level of traffic arrivals per direction. The aspect of fairness is applied in different ways dependant on factors including number of junction entrances, number of lanes per entrance, average traffic flow, peak traffic flow and popularity of destination; it can also be applied according to local politics or economics. The degree of fairness can be measured in terms of average queue length, maximum queue length, queue volume, average waiting time or maximum waiting time according to junction circumstance.

Inefficiency in a TDM system is introduced when a channel or route is temporally opened but there is no flow due to a lack of waiting traffic. Traffic light control systems have the same inefficiency and as a result vehicle detectors were introduced to reduce losses by identifying the most and least used approaches to a junction.

Traffic volumes and directions change according to the time of day and day of week, they may also be influenced by local events such as roadworks, incidents / accidents, mass attendance sports / social events and seasonally related activities (e.g. beach visits, January sales). In order to improve overall throughput efficiency, predictive and synchronised traffic light systems have been implemented.

## Controller Concept

This work identifies that the current methods of traffic detection and control are longer adequate for the optimal distribution of traffic under periods of congestion. The work identifies physical and logical changes that can be made to existing systems and employs a commercial traffic simulation package to show the degree of improvement that can be anticipated by the implementation of the proposed Available Forward Road Capacity (AFRC) algorithm for traffic signal control.

The AFRC controller concept is based on several mathematical principles, the most dominant of which is Queueing Theory. Existing systems use approach detectors to set the green demand. This is shown in Figure 5.1

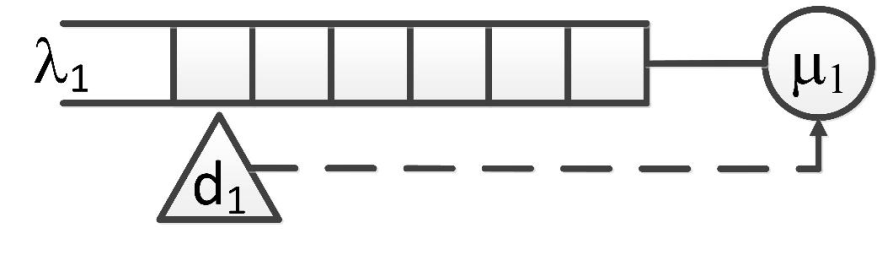


Figure .1: Approach Detection Control

Ideally a value for t is required such that λ vehicles are cleared by μ in t seconds however the value for λ is only known for the distance, d. While traffic is detected by the detector at d1, the time t will be reset :

Equ( ‑)

Therefore t is obtained from:

Equ( ‑)

The total value of t will be limited in a practical system by the minimum and maximum permissible green time. Equ( 5‑2) is correct while considering an isolated junction that is busy but not congested. However, in an urban network there is likely to be serial queueing that creates an output buffer, K, as depicted in Figure 5.2

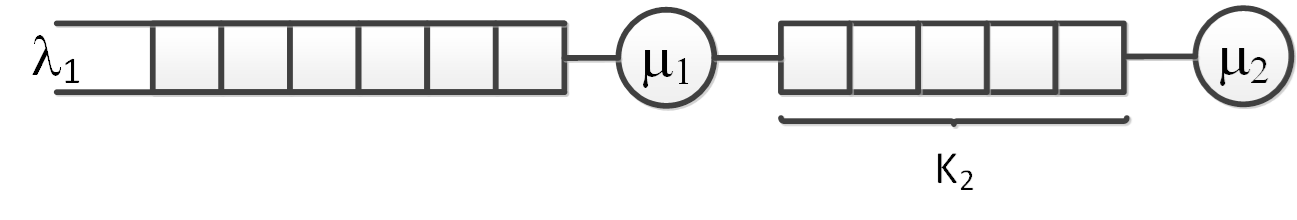


Figure .2: Serial queueing with an output buffer space

Existing traffic control systems estimate the vehicle capacity, K2 of a downstream link to determine the viability of forwarding traffic and adjust the junction timing according to the estimate. Without perfect estimation, the capacity will be either underestimated, causing inefficient operation due to early signal transition, or overestimated causing inefficient operation due to spillback blocking preventing practical control actions. In either case, the system operates below maximum capacity which needlessly extends queue and congestion duration. The physical and temporal occupation of any road system can be reduced by removing or reducing control system inefficiencies. The efficiencies will be reduced by improving local control and considering the road system in large but elastic sections that are self-organising and able to interact or act in isolation to improve the overall network throughput. Reducing the inefficiency will minimise the physical and temporal boundary constraints, reducing vehicle journey time, reducing pollution and improving pedestrian access.

The theoretical absolute maximum vehicular capacity of a roadway section is primarily determined by the physical dimensions of the roadway however the practical vehicular flow capacity is also affected by human decision and reaction. The decision process and, to some degree, the reaction process are influenced by legally defined protocol (e.g. “The Highway Code”) and by moral judgement. In an actively controlled traffic system, the moral obligations are minimised and the flow is principally determined by the active signalling. Modification of the signal system sequencing and timing therefore has the most significant impact on traffic flow at an isolated junction.

Traffic lights create a temporal segregation of traffic flows and, as a direct consequence, always reduce route capacity. Besides this intended result, there is an additional period of total time loss for every signal transition; no traffic flows in any direction for a period of time that is predicted to allow junction clearance before the conflicting flow commences.

Traffic flow at most junctions is commonly considered to be tidal; traffic controls are biased in favour of the presumed tidal direction to the potential detriment of other directions. The presumed tidal flow direction is determined in advance and is not modified to accommodate slow regional changes, sudden unanticipated events or the potential for pulse or dual direction tidal flows.

Where junctions are located within a moderate distance, the traffic output from one junction may lead directly to the traffic input to a subsequent junction. In this case the overall traffic flow will be influenced by the junctions and the interaction between the non-isolated junctions. In traffic terms, adjacent junctions may be isolated in one direction, both or neither and status is very likely to be temporal; junctions that are normally isolated may become non-isolated under heavy loading or vice versa.

Traffic may not be able to pass through a junction, regardless of signal indication, if a subsequent non-isolated junction is already experiencing congestion. As a consequence, congestion begins at a specific point on a route and builds backwards along the route path.

## AFRC Proposal

### General Description

The Available Forward Road Capacity (AFRC) system was conceived as a two stage process to reduce the effects of urban congestion. By considering the capability of traffic to exit from a junction in addition to measuring the traffic arriving, it is possible to improve the efficiency of the junction area as a whole rather than considering optimisation of individual direction timing.

The AFRC system is not intended to be used to control traffic flow unless congestion has been detected, the existing systems will manage sequencing and timing under normal conditions i.e. when traffic is flowing. The simple reason for this is that no improvements can be made to clear the congestion if there is no congestion to clear. Conversely, the AFRC system can improve traffic flows when the existing signalling system has been unable to prevent the onset of congestion.

The initial stage of AFRC is the prevention of a signal transition from red to green if it is not possible for traffic to move due to the exit being blocked by other traffic. Additionally, if an exit becomes blocked during a green period, the light should rapidly change from green to red to avoid a further unusable time delays. By assigning more red time to a traffic flow direction which cannot move, the green time can be allocated to an alternative direction with beneficial results. The initial stage of AFRC detection is shown in Figure 5.3. Detector d1a is used to determine the value for t as defined in Equ( 5‑2) however this is subject to the limitation of the size of the output buffer K2 shown in Figure 5.2.

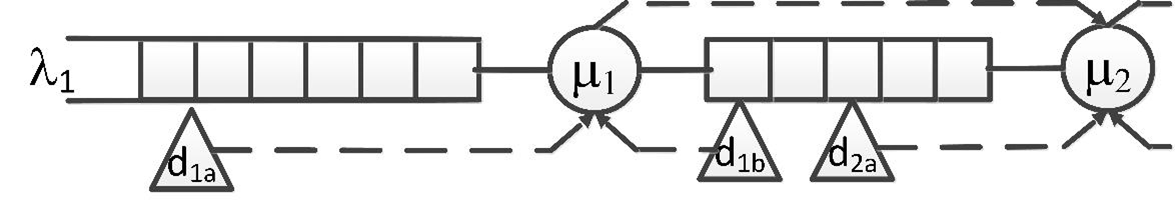


Figure .3: AFRC queueing control – stage 1

The inclusion of K2 however raises several problems. Firstly it is not possible to determine the capacity of K2 without obtaining :

* precise physical roadway dimensions
* exact dimensions (length) of the vehicles
* the gap between the vehicles

Second, the value of K2 sets a limit to the value of t as given by Equ( 5‑3)

Equ( 5‑3)

However the value of t to obtain K2 is not determined purely by μ as suggested by Equ( 5‑4)

Equ( 5‑4)

Each vehicle (λ) in the initial queue will progress into K2 at the lowest rate of either

* service rate μ
* vehicle capability
* driver preference
* vehicle capability of any vehicle in front (cellular automata)
* driver preference of any vehicle in front (cellular automata)
* delays from random events such as pedestrians

The final point regarding K; if the information required above is supplied and perfect for the junction under consideration then the solution will be specific to that junction and the solution will not fulfil the goal of being generic and universally applicable without tuning.

In order to avoid most, if not all of the problems of K2 estimation, the exit detector d1b is used to provide an absolute measurement of the available capacity of K2 in front of the signal to which it refers, the Available Forward Road Capacity. Detection of occupancy of the d1b detector will immediately clear the timer t. If t is cleared while the signal was showing green, the signal will change to red and a subsequent stage will become active. If t is cleared while the signal is showing red, then the signal will not enable this phase and no time will be lost to a flow that is not feasible.

In order to force a situation change, the second stage of AFRC is to generate an alert to request downstream signal changes to alleviate the local problem at source to aid flow in the most congested direction. The forward notification is shown in Figure 5.4

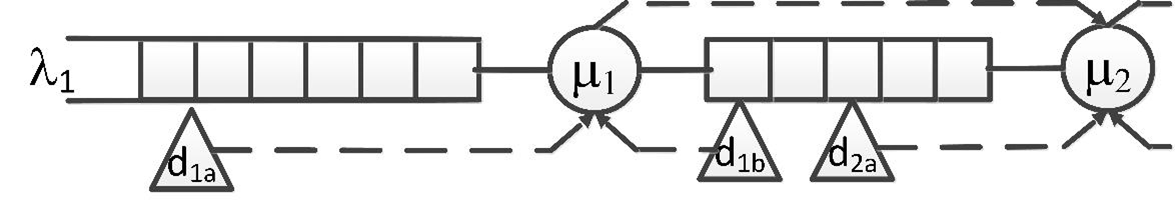


Figure .4: AFRC with forward notification

The system will be able to identify when an exit block has cleared in order to enable a flow to resume.

### Benefit of AFRC

Existing systems that rely on approach detectors can make a request for the allocation of green time but have limited ability to request red time. In contrast, an exit detection system would request red time but have no ability to request green time. While a control system can be designed around the AFRC system, AFRC is not proposed as a complete traffic light control, but as an addition to existing traffic control system. AFRC modifies the action of an existing green demand system when that system is not able to actively control traffic due to congestion build up i.e. there is stationary traffic on the stop line detector that extends the VA signal to maximum duration even though there is no actual flow.

AFRC is intended to adapt to actual prevailing traffic conditions and is not based on a predictive modelling approach. This means that AFRC will respond to actual conditions to determine the signalling strategy that is most likely to resolve overall congestion in the shortest possible time. In contrast to existing “green wave” solutions, the AFRC route decision, and indeed the next downstream junction, are not predefined or even necessarily persistent across consecutive signal changes.

AFRC affects each approach to each junction uniquely. Action is only taken when an approach is congested, all approaches are continually assessed and more than one approach can be declared to be congested.

Action is only taken when congestion is measured, there is no prediction and so the system remains effective even when there is variation from normal traffic flow e.g. an unexpectedly popular football match bringing additional traffic, weather conditions causing a change of flow timing, an accident or incident nearby causing a change in flow pattern. The continuous and immediate reaction of AFRC could not be matched by adaptive control systems that attempt to adjust over a number of cycles. Changes that are applied by adaptive systems during periods of exceptional circumstance will cause the predictive engine of the adaptive system to be incorrect when the exceptional circumstance is resolved. The adaptive system would therefore be ineffective during and after the event which will tend to extend the duration of the disturbance.

AFRC is only used when the specific approach to a junction is congested, for this reason the AFRC system does not require any tuning data or tuning period in advance. The only requirements will be to know the exit detector associated with each signal phase and the means to communicate to the next downstream junction associated with an exit.

There are a number of points that contradict the accepted design constraints for regular signal systems. These could be considered to be major issues that preclude the use of AFRC however they are addressed as follows:

There should be a minimum amount of time for a green indication to be shown. AFRC will change the signal from green in immediate response to the exit being detected as congested which could be in less time than the minimum. This is not considered to be a problem as traffic will not be able to move regardless of the signal indication. This will not cause confusion or problematic traffic movements as the situation will be clear to the road user. The function of AFRC under these conditions will be the same as existing “Yellow Box” junctions, vehicles will not be permitted to enter the junction unless the exit is clear however AFRC will have the advantage of moving the traffic control sequence on to the next stage allowing traffic flows in uncongested directions. This reduces the amount of time that represents total loss of flow.

AFRC may skip green phases showing extended red indication to specific directions which may cause driver frustration and prompt civil disobedience of proceeding against the indicated red signal. This is again not considered to be a problem as the green phase will only be missed if the intended exit is blocked. Drivers will be no more frustrated by the indicated signal than they would by the congestion itself. The possibility of civil disobedience will be minimal as the stop signal will reflect the condition of being unable to proceed rather than indicating an instruction not to proceed when there is a clear option. It is thought that there would actually be a lower incidence of red light running if AFRC is making maximum use of the available road space.

There is a minimum of two seconds duration for the yellow (amber) signal. AFRC should be able to theoretically reduce the yellow time to zero because the decision to force a change to red is based on the traffic already being stationary, there should be no need for a warning that the signal is going to change to red in this circumstance. This again reduces lost time by enabling an earlier transition to green for conflicting flow. This reduction was implemented in later versions of the AFRC control system.

### Changes Required To Accommodate Practical AFRC

A new detection signal for each junction exit is required to enable AFRC. The detector could be any type discussed in section 2.3.3 provided that it can identify stationary vehicle presence, Doppler shift RADAR is therefore of limited use. The reason for this detector is to provide absolute feedback on exit status in preference to derived / assumed conditions.

It is possible to assume the level of occupancy of a junction exit based on the entry stop line detector if the traffic signal is showing green, i.e. if the signal is green and traffic is passing the stop line detector at the junction entrance then it should be a reasonable assumption that the exit is clear. However there are limitations to this assumption, for example there is no way to confirm that all possible exits are clear, it is possible that only one is clear. It is not possible to confirm the status of an exit under a red signal when there is no flow to detect and therefore the exit status cannot be confirmed before a change from red to green with the potential for lost time. The proposed system must detect when a junction is blocked but it must also detect when the block is cleared, this again would not be possible with the existing approach detectors alone.

There will be a time delay between the exit of a junction becoming blocked and the flow ceasing at or before the stop line; this is time that could be allocated to alternative flows.

If the traffic is detected as not flowing into a junction, this does not confirm absolutely that the exit is blocked. An indication that there is no flow in to a junction when the exit is clear could indicate an incident / accident within the junction that could be used as a notification to Police or Traffic Management groups.

The exit detector could be mounted above the existing signal head and use existing wiring or wireless connection to the controller.

The second change required to implement a practical AFRC solution would be a change to the junction controller to accommodate the new detector signal, a signal from an upstream controller and to be able to transmit a signal to the next downstream controller. The software changes required, storage space and processing time are minimal and will not require additional processing hardware beyond the hardware that currently exists for each junction signal system.

The final requirement would be for a communication link to each signal controller to allow signalling to be communicated between each upstream – downstream pair. This link would not need to be dedicated physical point to point, communication could be achieved over existing lines used for monitoring by the application of a logical point to point communication channel.

Junctions could be integrated into an AFRC system as required, junctions would not all need to be connected immediately. Overall cost and inconvenience of implementing a practical system would be minimal.

### Future Need for Traffic Control

There will still be a need for some form of independent arbitration system while the demand for road space exceeds the actual space available. Drivers / vehicles cannot decide on their own personal and route priority above other road users otherwise every vehicle will prioritise itself above all others and the will be an impasse for every vehicle at each junction. With no priority scheme, or an even level scheme, the volume of traffic approaching a junction from each direction will be different and some form of load balancing will need to be applied otherwise the busiest route will become totally congested while the minimal use routes will clear rapidly. The load sharing concept is no different to the fair allocation processes currently employed by traffic signal systems; the decision must be independent and based around the physical constraints of the junction, route and traffic volume. Regardless of the physical appearance or method of communication, it is considered unlikely that the “traffic light” will be permanently removed.

While the demand for physical transportation persists it is almost certain that there will be conflicting directional requirements that will lead to congestion and the need for traffic control systems. The AFRC algorithm will remain a suitable approach to minimise congestion for all traffic systems and may be adaptable to other purposes including Production Engineering and Communication Systems Engineering.

## Algorithm Design

### Introduction

The concept design is to measure actual vehicle locations to determine if junction exits are blocked and utilise this measurement to influence the sequencing and timing decision process for existing traffic light controllers. Junction exits may become blocked due to unexpected events such as accidents or, more commonly, due to spillback queueing from downstream traffic control. In either case there is no realistic possibility of traffic moving through the blocked exit and therefore no reason to direct traffic flow towards the blocked exit. This serves no useful purpose and at the same time the signal action generates and extreme level of “lost junction time” by preventing possible flow in conflicting directions that could assist in clearing the overall congestion period.

The new control therefore has to:

* Detect blocking
* Prevent flow towards blocked exit

However this simple approach does not resolve the problem and, as early experiments proved, can promote gridlock. To prevent this, the new control also has to:

* Identify sequences that do not include the blocked exit
* Promote clearance of the blocking
* Test for blockage clearing

To maximise overall throughput efficiency controller must:

* Consider all flow directions equally regardless of presumed “normal” flow patterns
* Minimise lost time due to signal transitions
* Prioritise traffic flow that leaves the congested area

In order to provably achieve these design criteria, the design of the control logic has been conducted in five main stages:

1. only local junction implementation
2. as 1) but with forced downstream change ignoring downstream sequence preference and timing
3. as 2) but implementing sequential light phases as opposed to simple ns / ew flow
4. as 3) but including upstream information in the sequence / timing decision process and requesting rather than forcing downstream changes
5. as 4) but reducing lost time due to transitions

The initial traffic light controller designed as part of this work was constructed from a simple VA system shown as a flow chart in

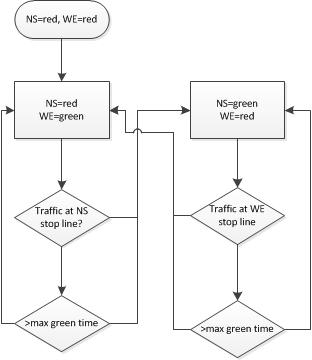


Figure .5: Simple VA Control

The simple VA was then developed to include a minimum green time and incorporate the exit detector. The flow diagram for this stage is shown in

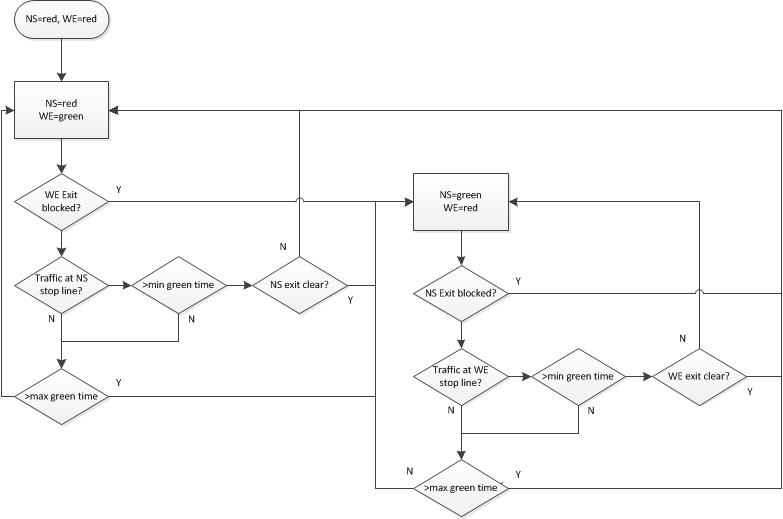


Figure .6: VA tls with Exit Detection

The sequence of was coded in Python and used as the initial control process for the traffic lights in the Aimsun roadway simulation shown in . Traffic was simulated using fixed traffic light timing and using the exit detector.

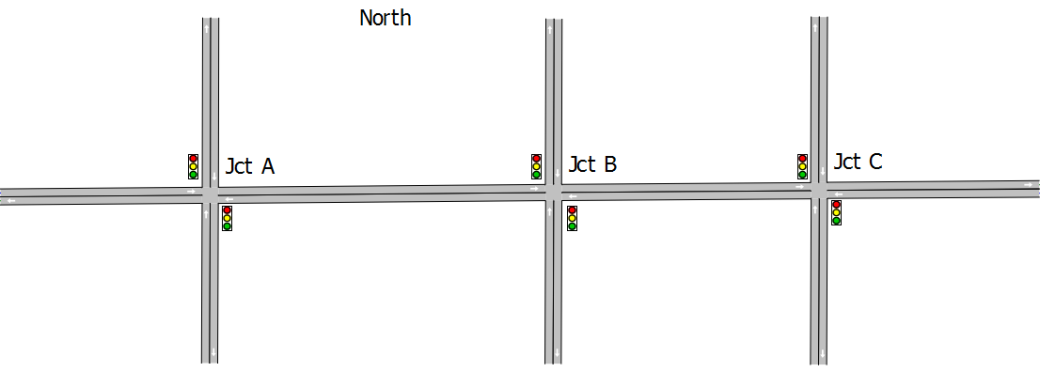


Figure .7: Aimsun Simulated Roadway

The use of this layout is discussed further in the Testing section of this chapter under “layout version 2”

### Testing

A very important aspect of any testing regime is to ensure that the tests provide results that are representative or indicative of the actual use case. In traffic simulation either a real situation or a representative approximation needs to be considered. While it is simple to suggest that an actual situation is simulated with a real known outcome, this concept can be completely misleading.

It is accepted that experiments should be repeated a significant number of times to obtain an average of the results, standard deviation and a confidence level. Using the results obtained from a single situation with a single outcome defies this logic. Real traffic presented to a real road system has variability in physical and temporal presentation; the same drivers do not respond consistently, present the same vehicles, in the same order and at the same time on any given day. By accepting that there will be significant variation in any or all of the input criteria and accepting the principles of Chaos Theory it must be accepted that the outcome will not be constant or repeatable; the input is only an approximation and this will not provide an approximation of the future outcome. The possible influence of external forces and actors compounds this disposition; this has been discussed several times and the impact shown, notably in section 4.5. The effect of external influences ensures that the entire scenario is never a constant. It can therefore be deduced that the results obtained from a single scenario that contains so many variabilities cannot be taken as mathematical or physical certainty. It should not be assumed that the same outcome would result even if it were possible to have an apparently identical set of initial conditions (which it is not); any outcome could easily be an outlier.

It is well known that traffic volumes and flows vary over time on any given route, a significant amount of the SCOOT process is dedicated to adapting to this change as discussed in section 2.3.4.5. Given that this is an accepted state, and that there are random weather and cyclic seasonal variations it is not possible to repeat a traffic experiment with any degree of outcome certainty other than a gross, macro or bulk similarity. The micro level outcome can vary over a very wide range.

As discussed in the simulation topic in section 2.4.7, the output of the psycho-physical system is a probability function based on the combined probability functions of the interacting components. Repeating the same simulation while varying the component probability functions could provide a degree of confidence of an outcome assuming that a) the inputs remain constant and b) there is a reasonable level of confidence in the capability of the simulation package. In order to create a change in the probability functions, traffic simulation systems such as Aimsun use a seed value which is exposed to the simulation configuration as an initial condition. In order to generate variable results the seed value is chosen as a random number before each simulation. However this is not the goal of the simulation tests that are being undertaken for the purpose of this thesis.

The goal of these simulations is to examine the outcome of a scenario when different control methods are used. In direct contradiction of a normal traffic simulation experiment, the testing will generate the same results each time the simulation completes meaning that the average of any number of simulations will be the same as any individual simulation; there will be zero standard deviation. To achieve this objective, the seed value will be fixed and remain the same for all simulations. This is a concept of minimum change and can only realistically be achieved in a simulated environment where there is total control over every aspect of the experiment.

The major components of a traffic simulation scenario are listed below and are subsequently expanded upon:

* Vehicle (human) responses
* Initial conditions
* Road / Route layout
* Traffic profile
* Traffic control system

The subject of this thesis is to develop and test the ability of new and unique traffic control system algorithms to reduce overall congestion queueing and duration. In order to achieve this, different control systems were compared to fixed timing and vehicle actuated systems in order to determine if improvements are possible. Testing is undertaken against road systems and traffic profiles that are progressively more complex to address the limitations found in systems from the literature review and to prove that the AFRC control algorithm offers a general solution.

Tests were undertaken using six different road layouts and six versions of AFRC controller. The traffic profiles are adapted to the layout under test and to represent more realistic conditions.The AFRC developments are progressive in order to resolve the more complex layouts and traffic flows however they are totally retrospectively compatible, each later system could be applied to the simpler earlier layouts.

#### Vehicle (human) responses.

As described earlier, this is the most complex element of the traffic simulator as it is based on probability functions of physical and psychological response. This is also the easiest element to control from a repeatability perspective. Setting the seed value to a known fixed value at the start of each test will ensure an identical result for each simulation run assuming that there are no other changes to the scenario. For these experiments the seed value was set to 10000.

#### Initial Conditions

The simulation will begin with a calculation of a best path from an Origin to a Destination (OD path) followed by traffic entering the system at the Origin and following the OD path to the Destination. The OD path is periodically recalculated based on current traffic flow and volume. At the start of the simulation there will be no traffic inside the system and hence no useful control action. In order to make the entire simulation result useful the simulation can have a preload time which sets the vehicles to the position that they would be in after a given amount of time. This is achieved by actually running the simulation at the fastest rate and without any graphical update until the preload time completes. Any path recalculation is also completed during the preload. For these experiments the preload time was set for 15 minutes and the route recalculation set to 6 minutes.

#### Road / route layouts

The layout sets the permissible locations for vehicles to occupy but it also sets travel boundaries and the rules to follow, e.g. driving on left or right, road speed, lane direction, junction priorities and so on.

Layout version 0 was a simple crossroad junction with NS and EW paths as depicted in Figure 2.1 and Table 2.2. It was isolated and not representative of any particular flow. The purpose of the junction was to allow the creation of timed and VA traffic control signals with stop line detection. The testing and results from this layout are not included due to their rudimentary nature.

Layout version 1

The initial motivation for this project was a group of junctions in Derby. The thesis rapidly expanded to consider the general problem rather than the one very specific situation however before completely discounting the “Derby junctions”, an attempt was made to simplify the road layout into a functionally equivalent model. Diagrammatically removing corners and changing entrance and exit compass headings provides a better understanding of the problem. The simplification, which is also not to scale, is shown in Figure 5.8.

The diagram, coupled with a small degree of local knowledge, shows how this route will always be congested during busy periods due to physical layout. As a result, even though some benefits are available, the route is not a suitable candidate for large scale testing of traffic light control systems.

As a brief problem description:

Four lanes of traffic enter from the bottom of the diagram at 40mph. Local knowledge, observation and traffic overlay data from sources including Google maps [119] and Bing [120] taken during peak hours suggest that a significant proportion of the traffic takes the left fork. Traffic attempting to leave needs to be in either lane 1 or lane 2 on the left. Any queueing traffic in lane 1 or 2 restricts the lane change option of traffic in lane 3, reducing the capacity of the major “inner ring road” route.

The first junction (Queen Street) is traffic light controlled and, from observation, appears biased towards traffic entering the main route, this is supported by the fact that the entrance serves public transport both entering and crossing the main traffic flow. The junction is also a pedestrian crossing.

The restricted flow through these lights causes traffic to queue back to the four lane major route, causing the lane restriction described previously. Progressing up the diagram, and therefore downstream from the lights, the speed limit reduces to 30mph, a change which has no impact during congestion periods but may promote the onset of queueing during the transition from free flowing to congested.

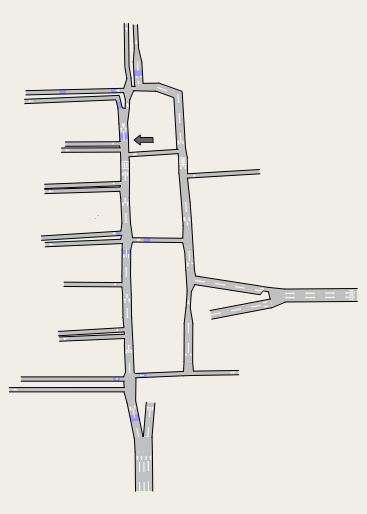


Figure .8: Diagrammatically Simplified “Derby Junctions”

The next junction is give way line controlled. Some traffic leaves the main route at this point towards the West of Derby but other traffic enters the route towards the residential areas and main exit routes from Derby to the North. There are two major multi-storey carparks in this area to add to the flow. Observation suggests that more traffic enters the considered route than leaves at this point. This junction also offers a potential “time advantage” to avoid some of the queueing traffic leading to the Queen Street traffic lights. Entering at this point offers logical and mathematical advantages; being downstream of Queen Street traffic lights means that the major flow is interrupted by the lights providing a frequent 5 second break in flow on the major route. Further downstream the traffic density increases due to the speed limit restriction and lane narrowing to accommodate a section of bus lane with priority traffic light, which is again a pedestrian crossing with random change requests. Queueing for this section leads directly into the previous traffic light controlled junction. A one way street enters the main route. Even though the traffic entering at this point is normally minimal, traffic can only enter the flow and not leave.

The subsequent junction is only a short distance at Lodge Lane. The lights are three way with the one way major route entrance, a two way entry/exit and a one way circulatory route entrance. All entrances to the junction are busy especially during rush hours. The circulatory entrance has the potential to cause spill back into traffic in the complete opposite direction to the direction under consideration with the potential to become a part of a gridlock state. Traffic can leave to the West / South from this junction however the time benefits of leaving the route at an earlier stage and the time penalties of using this route mean that minimal traffic leaves using this route. With minimal traffic leaving, and traffic entering from two directions, there is again an increase in traffic density. There are other one and two way entrances to the roadway, a popular petrol station that is used as a cut-through in both directions and a pedestrian crossing (indicated by the arrow) before the final junction on the route. The smaller junctions are downstream from the pedestrian crossing which offers a tactical advantage for traffic entering from the west of Derby. The final junction separates the two lanes, one generally towards the North-West and one to the North. The Northbound traffic is almost invariably a greater volume as in provides access to a greater number of destinations, many of which have very limited alternative routes. The Northbound traffic has to contend for the junction with traffic from the West and traffic from the North. Circulatory traffic can, and frequently does, spill back form the junction; as few as 5 cars / light vans can block the junction. Traffic density and delay on the Northbound route and the Southbound circulatory route peak at this point due to the flow being restricted to using only the single right-hand lane.

In summary, two lanes of traffic arrive at 40mph to a three way temporal restriction where traffic is added and the speed limit reduced by 25%. Traffic proceeds to a further three-way temporal restriction, further traffic is added and the available road capacity is reduced by 50%. Finally traffic leaves through a further temporal restriction which has a high probability of spill back blocking.

While AFRC could potentially improve the flow of traffic passing through the three main traffic light controlled junctions, the main effects of road layout and funnelling cannot be directly resolved. Poor design and layout restriction due to Listed Buildings mean that the options for this area are limited although it would almost certainly benefit from the removal of active control in all directions on Duffield Road.

This layout has very specific problems and cannot represent a generic situation. It was not considered further for testing in its entirety but led to the development of the next layout.

Layout version 2

The detail points to the “Derby Junctions layout” are only of value if real traffic flow data collected over a significant period of time was available and a solution was required for the one specific junction at a particular time of a given day. The traffic flow is not readily available and the research outcome would be limited. The junction concept was however developed through simplification, this removed many of the specific features that make this section of road unique and therefore should identify a more general solution. The result of the simplification is the three sequential traffic control points as shown in the Figure5.9 layout which was used for some of the early proof of concept design. This layout also uncovered some of the flaws in a basic exit detector concept and the method of control system programming.

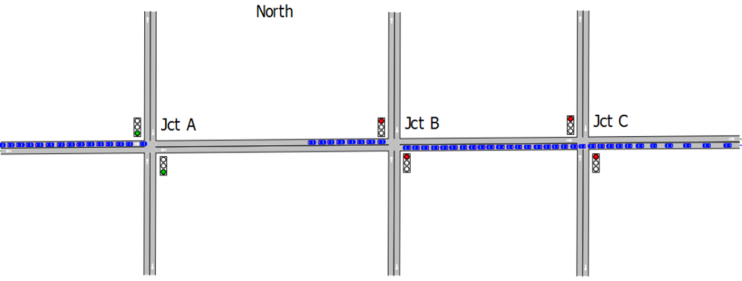


Figure .9: Simple Three Junction Layout

An hour of traffic flow was simulated with flows from West and East at the maximum road capacity ignoring junctions, each North and South junction entrance was presented with a minimal 10% capacity. During the initial time period, traffic flow was as predicted with lights changing as anticipated however two problems became apparent as the test continued.

Firstly, a gridlock situation occurred after a short period. Two causes for this were that the simulated vehicles would move into the junction during the green phase before the signals had changed to red regardless of the exit being blocked. Having entered the junction, they were a physical obstacle that prevented other vehicles from completely passing through the junction. However the other vehicles still entered the junction but could not leave. Each path became blocked by the conflicting traffic flow which quickly caused a tailback to block the preceding junction creating a circular gridlock. This situation is permitted by default in the roadway design of Aimsun and was resolved by making each junction a “yellow box junction” where traffic will not enter a junction unless it is also able to exit the junction. The second cause was that traffic blocking the East exit of a junction (e.g. Jct B) was detected and caused a change in the traffic control phase (by design) even though the West exit was clear. Traffic approaching the West entrance was therefore blocked and the queue tailed back to the traffic light to the West (e.g. Jct A). When this blocked the exit, the traffic light changed to red and blocked the East and West flow preventing the original junction (Jct B) from clearing. Again there were two problems identified, first the simple NS and EW control system means that traffic flows always directly affect the opposite flow direction and second the junction signals operate in complete isolation; upstream junctions are influenced by downstream control but do not exert any feedback.

The second problem was that traffic flows were higher when the controllers for Jct A, B and C were set as fixed time rather than VA. This was not initially anticipated as it implies that there is no advantage to using VA systems. The research presented in section 2.3.4.1 however shows that fixed timing with minimum signal transitions results in higher throughput than VA systems when the junctions are over saturated. This combines with an anomaly caused by synchronisation in the simulation timing. Within the simulation, each traffic light controller has the same initial start time and the same reference ”clock”. This led to the signal phase of each junction being synchronised which results in the three junctions appearing as a single entity and allowing a rudimentary and unplanned “green wave” flow. As a method of proof, the signals were given a timing offset

The layout provided initial experimental results has been used to support the explanation of downstream blocking and congestion pockets however the lack of interaction through alternative routes made this unrepresentative of a realistic general case. Even though the testing was limited, the results were invaluable in developing the AFRC system.

This layout version was discounted from the formal simulation results primarily because it fails to adequately represent the original junction. Any traffic data applied cannot provide a realistic solution to provide a comparison. The layout does not provide significant interaction opportunities and so lacks any degree of realism. For example, a gridlock situation cannot exist without very precise flow control definition; if the gridlock is created in this way, then it is highly contrived and easy to resolve.

The control system was programmed for each signal individually. Two of the design goals were that each controller would be identical and that none of the controllers would be tuned to the requirements of a specific junction. With only three signals to control, this was not a large task but it made the code base very large and led to occasional errors due to mistyping.

The diagrams in Figure 5.10 and Figure 5.11 are included at this stage to clarify the general locations of the entrance and exit detectors. It should be clear that the stop line detectors are reasonably accurately placed; this is a requirement in a practical road network and is accurately defined by documents such as [30], [32] and [97]. The exit detectors are far less repeatedly placed. This is deliberate and intended to show the lack of precision required during installation.

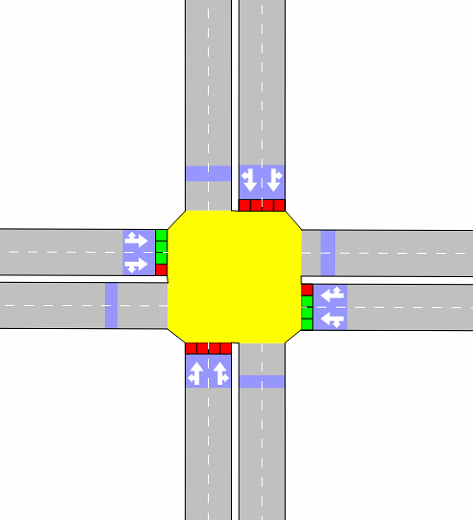


Figure .10: Junction Layout With Detectors

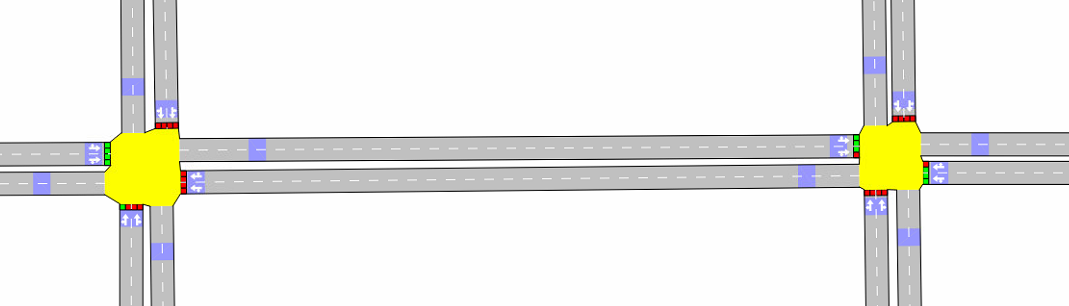


Figure .11: Typical Upstream/Downstream Link

Layout version 3

To resolve the limitations of previous versions and offer greater options for experimentation, a “Manhattan” or “Milton Keynes” style grid was defined as shown in Figure5.12. Having noted the failings of version 2, the concept was to develop a system where traffic would flow to and from opposing corners providing maximum vehicle interaction. The small circles at each corner are the simulated traffic Origin and Destination points. Each junction is a yellow-box traffic light and operates almost identically with shared code. The code created for this version considers each approach direction individually as defined in Table 4.1.

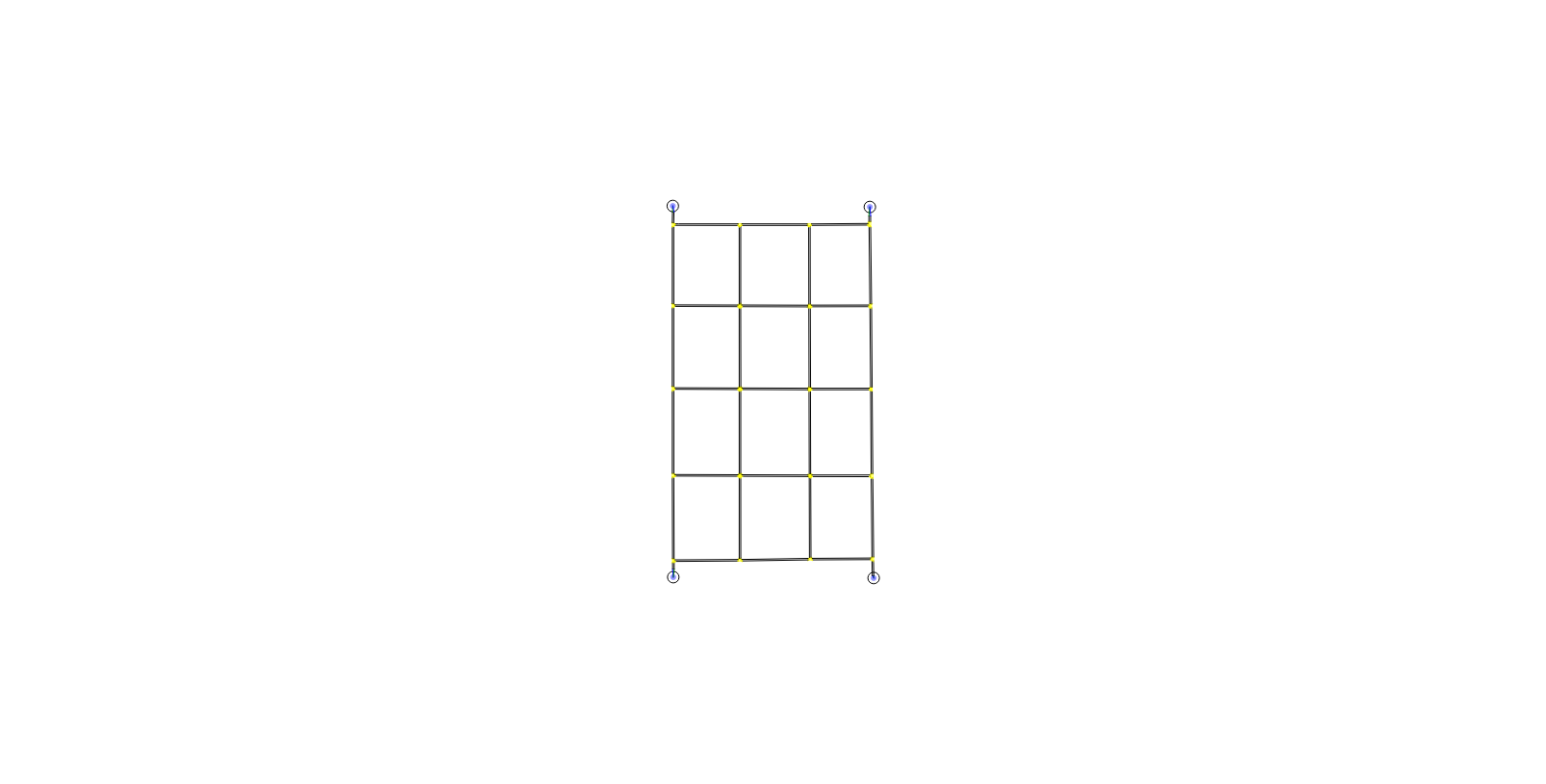


Figure .12: 4x5 Traffic Grid

The testing and results are explained further in Chapter 6:. This layout was generally successful in the development of the AFRC algorithm and in visualising the difference between timed, VA and AFRC systems but limitations were observed that prevented this from being considered as a final solution.

One of the first limitations of the 4x5 grid is that the edge nodes have three approaches while core nodes have four. If the code and sequencing are applied equally across all junctions this results in wasted sequence intervals where edge junctions do not pass any traffic. In order to avoid this, the timing of each edge node was modified from the core timing. For example, the midpoint node on the west side does not have an approach that equates to phase 1, therefore phase 1 timing is either wasted or needs to be shared to the other phases. For the same node, the missing approach also affects other phases; phase 3 no longer has a major exit. The timing of the edge nodes is different for each side of the grid and is acute at the corner junctions, where traffic enters and leaves. While this modification is consistently applied to each of the control systems, it still fails to meet the requirement of each node being identical.

The major limitation however is in the action of the entry and exit traffic lights which restrict the ability to inject traffic at volumes that significantly exceed the core network capacity. Excess traffic injection is possible when the previous modification is applied due to there being fewer phases and therefore fewer transitions and less lost-time but there is still a restriction that leaves traffic queueing outside of the system waiting to enter. This is the same effect as gating discussed in 2.3.4.5 and 3.4.3 limiting the requirement for the controller to manage all traffic and artificially showing a control system improvement by moving traffic queues outside of the area of interest; the delays still exist but are no longer observed or considered.

Layout version 4

The differences with the edge node phasing was considered to be the major failing of the 4x5 grid and as a result the 7x6 grid was developed. The decision to change the name orientation was made to be more consistent with the junction numbering used in the code.

The 7x6 grid allows all of the junctions to have identical phasing as all junctions have the same number of approaches even though some have an infinite destination. The action of the traffic light controller for VA / AFRC means that the unused approaches are only activated when there is little or no traffic at any approach and the control action reverts to sequentially timed.



Figure .13: 7x6 Grid, All Junctions Have Four Approaches

The version 4 layout also allows traffic to take a route which is not the most direct; under congested conditions a vehicle could take a route which is physically longer but less congested.

The position of the Origin and Destination points mean that the network is still being treated as a transitory route i.e. traffic enters and leaves the network at extremities without considering what may happen beyond the limits.

Layout version 5

The version 5 layout recognises that traffic rarely traverses an area completely; the Origin and Destination are more likely to be within the network rather than at the extreme edge of it. This would be more representative of a city region where the majority of people work and live inside the civic area. It this case the Origin and Destination points can be considered to be residential estates, industrial estates or mass employment car-parking. Moving the Origin and Destination points slightly in from the edges and adjusting the configuration allows traffic a choice of direction to enter and leave the network effectively doubling the input rate capacity. As the entry and exit is no longer controlled by a traffic light, traffic flow is not restricted in the same way as previously. The change also means that traffic can pass completely around the source and destination points rather than passing though. The choice of traffic Origin and Destination when combined with tidal traffic volume enables the effect of traffic convergence and divergence to be observed.

This layout is intended to provide significantly more realistic flow rate and direction and is referred to as the “realflow” grid.

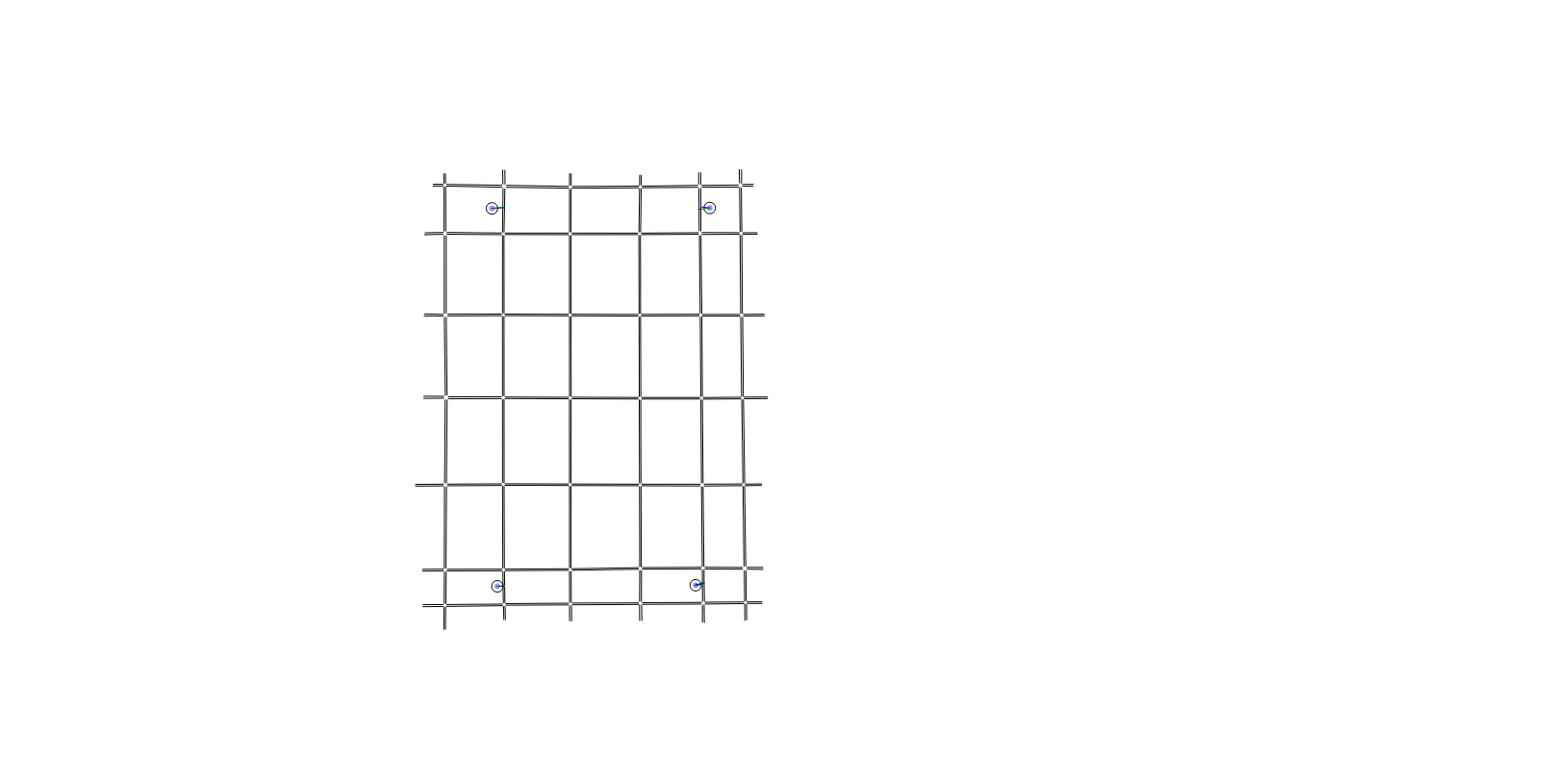


Figure .14: 7x6 “Realflow” Grid

Layout version 6

During early testing it became apparent that although the traffic Origin/Destination had been moved away from the edge of the network in version 5, traffic was entering and initially travelling in a single direction e.g. North or South but not both. Similarly, traffic could only leave the network when travelling in a specific direction. This means that traffic access is still via a single specific traffic light which would cause an input and output delay. It was also noted that in the version 5 layout, traffic that was waiting to enter the network from the Origin was not able to activate the exit detector for the route traffic controller and this was preventing a desirable signal change. This is especially problematic with the situation shown in Figure 5.15 where there is no traffic waiting at any of the other junction entrances that would be significantly delayed but there is clearly traffic being prevented from entering the system. This is the exact type of problem that AFRC is proposed to solve; the minimal flow of vehicles at the junctions is correctly activating the VA signals but this is causing longer delays for a greater number of other drivers.

The layout version 6 is identical to layout version 5 with the subtle difference that traffic can enter and leave in both directions and the exit detector is relocated on roadways that include a Destination as shown in Figure 5.16.

|  |  |
| --- | --- |
| Figure .: Traffic Queueing to Enter From Commercial Area | Figure .: Modified Commercial Area Link with Additional Detectors |

#### Traffic Profile

The traffic profile was developed in conjunction with the road layout and is intended to either make most use of the layout or to represent some general form of traffic flow. The traffic profiles are included with each experiment and an example is shown for clarity in Table 5.1. The id:name is the Aimsun allocated unique id number followed by the user defined name. In this case it is the ordinal quadrant of the Origin/Destination point. The top row of the table shows the Destination, the first column shows the Origin or source and the intersecting number shows the number of vehicles per defined flow duration. In order to maintain clarity, the duration of each flow was defined for these experiments as one hour which means the figure shown is in veh/hr. In these experiments the only vehicles were cars but a profile could be built for different classes of vehicle such as HGV. As explained in the introduction of section 5.4.2, the intended purpose of this simulation testing is to directly compare the capability of control algorithms rather than simulate a system to identify a solution. For this reason, there is no benefit to include or exclude multiple vehicle classes provided that the input is consistently applied. Given no benefit to include extra vehicle classes while their inclusion could simultaneously introduce a potential for error, the decision to use only a single class is considered correct.

Table .1: Example Traffic Flow Profile

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id:name | 775: South-East | 783: North-West | 1446: South-West | 1449: North-East | Total |
| 775: South-East | 0 | 100 | 0 | 0 | 100 |
| 783: North-West | 100 | 0 | 0 | 0 | 100 |
| 1446: South-West | 0 | 0 | 0 | 100 | 100 |
| 1449: North-East | 0 | 0 | 100 | 0 | 100 |
| Total | 100 | 100 | 100 | 100 | 400 |

A traffic demand is then constructed by applying the flow profile to the layout at a particular time, for a duration of time and with a given control process. The traffic demand can include many profiles simultaneously and/or sequentially over a period of time to create a piecewise input profile. It is also possible to use a pre-recorded real or simulated input but again, this creates complication without significant benefit.

Table .2: Example Traffic Demand

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Traffic Flow  (hh:mm) | 00:00-01:00 | 01:00-02:00 | 02:00-03:00 | 03:00-04:00 | 04:00-05:00 |
| Busy | Normal | Quiet | Quiet | Quiet |

The traffic demand shown in Table 5.2 is an example used in the experiments conducted in Chapter 6:, the time is 24 hour format and specifies the start and finish time of the traffic flow sections. For clarity and ease of use, the experiment time begins at 00:00 in preference to some realistic time e.g. 07:45, this choice means that the current experiment time can be simply read from the results tables and graphs rather than having to compare to the start time. As a consequence of the decision to create traffic flows with one hour duration, some of the flow sections are repeated to create an extended section. The flow name is defined by the user for each flow; names were chosen that indicate the level of traffic presented during the time period

In a real network traffic is very unlikely to flow across a network in purely conflicting directions. It would be more realistic for vehicles to flow from numerous (residential) areas to a few central (commercial or industrial) areas at one time of day and then return from the commercial area to the residential areas at a different time of day. The objective is to simulate travelling to work and then home again later. This was one of the objectives for the “realflow network” in Figure 5.14. Despite the similarity of layout between versions 4, 5 and 6, the traffic flows were completely redefined for each experiment.

The region traffic flow is scaled to represent areas in a city where the primary or most common use can be defined. This is totally hypothetical but was intended to given different traffic flow profiles across the entire simulated area with a reasonable justification of the vehicle numbers, directions and times:

South-East. Commercial / industrial centre, this is the main target destination for traffic during the early day rush-hour period. In some experiments it is also the main source of traffic during the late day rush-hour. This is intended to represent a large volume of people travelling to and from a place of work.

North-West. Residential 1, a high density large residential area where a high proportion of the workforce live. As a modified demographic, this area has a higher proportion of young families with one car but two drivers. The implication is that there will be a greater number of people are taken to work and the car then taken home again rather than being left in the commercial area for the duration of the day i.e. the traffic flow will be both directions in the same period.

North-East. Residential 2, a lower density residential area but with a higher proportion of vehicles that will be parked in the commercial area for the working day. Mainly one way traffic flow in any period.

South-West. Residential 3, a small residential area. Mainly higher managerial / retired people, not generally active during the main rush hour but likely to make journeys to other areas during the off peak hours.

The flow and demand are given in Chapter 6: for each experiment but they generally include a stage of very busy (overloading capacity), normal and very quiet. The intention of the quiet period is to determine the duration of the residual traffic flow.

#### Traffic Control System

The traffic control system is in three parts: Timed, VA and AFRC. The algorithms relating to Timed, VA and AFRC are given in Appendix 1: however the main concepts are given in Algorithm 1- . These are shown in reverse order of complexity to illustrate that disabling the exit detector of the AFRC will result in VA operation and disabling the stop line detector of VA will result in Timed operation. The controller code uses flags to determine if the exit and stop line detector values should be used (full AFRC), only the stop line detector value should be used (VA) or if neither should be used (Timed).

In order to make minimal changes for each stage of testing the layout, demand and controllers are not changed other than to logically disable the detectors depending on requirement. No hybrid testing was conducted; all controllers are either AFRC, VA or Timed in each test section. This approach works well as it will also prove the response of a practical controller if each function is selected, manually deselected or failed. The response of a detector in a failed (not sending) condition is identical to the function being disabled (ignored) or there being no traffic to detect. Because the controller is able to function totally autonomously with no failover time or input, it will continue to function under all circumstances other than total power loss or cpu lockup and will progressively fail towards the least dynamic but highest throughput solution of minimal change timed only.

|  |
| --- |
| If showing red:  If exit is blocked  Do not change to green  Request downstream change  (other AFRC specific Changes)  If exit is not blocked change as VA demand  If showing green:  If exit is blocked  Change to red  Request downstream change  (other AFRC specific Changes)  If exit is not blocked change as VA demand  Repeat |

Algorithm : AFRC

|  |
| --- |
| If showing green:  If maximum green time is reached  Change to red (next sequence)  If minimum time is reached  If traffic is detected at stop line  Extend green time  Traffic not detected at stop line  If traffic at conflicting stop line  Change to red (next sequence)  Repeat |

Algorithm : VA

|  |
| --- |
| If showing green:  If maximum green time is reached  Change to red (next sequence)  Repeat |

Algorithm :Timed

As previously stated, the AFRC system will only provide improved traffic flow under saturated traffic conditions when there is a spillback of traffic causing undesirable interaction with upstream controllers or conversely undesirable interruption to capability caused by downstream controllers. Because the level of spillback congestion is actually being measured there is no need for estimation of capacity (as used in many systems including [17]) which is likely to be prone to error due, amongst other reasons, to Human interactions (section 2.4.8), chaos theory (section 2.4.1) and game theory (section 2.4.6). Also as a result of the dynamic measurement of spillback, only those junctions directly affected will cause a change of sequencing and timing rather than the SCOOT approach of retiming a region based on one nominated main junction (section 2.3.4.5). Finally, spillback measurement and inter-junction ad-hoc communication mean that all affected junctions form relationships as needed, the coordination is totally elastic, including as many (or few) junction controllers as necessary, in the most suitable order and for a duration of cycles ranging from 0 to ∞.

The AFRC control systems were developed throughout the experimentation phase. The major changes are noted with the experiment that they relate to. All controller code was written in Python and applied using the Aimsun API.

The controller file is the main reason for a new experiment being declared as it determines the response of the system. New controllers were applied to the existing layouts until unsurmountable conceptual or logical limitations of the layout were identified. The functional aspects of the controller files are totally interchangeable across each layout however the implementation contains a listing of associated detectors and upstream / downstream junction. Changes in the layout may change the junction and detector numbering, changing the configuration section and invalidating how the controller is applied. The numbering changes required to apply a more advanced controller to a less challenging retrospective layout were not considered to offer developmental advantage.

# Simulations and Results

## Section Outline

In order to test the AFRC algorithm a number of experiments were conducted and the results obtained. The results presented summarise the simulations that were undertaken; note that some experiments and the results from some experiments were excluded when the experiment was considered to be procedurally incorrect.

Reasons for exclusion include:

* experiments repeated to confirm the results obtained;
* experiments that displayed minimal change in results from previous experiments;
* experiments conducted with coding faults that were immediately corrected, these experiments were repeated and the results are reported where relevant.

As a development process, all of the results were considered and no experiments are excluded simply because the results were considered unfavourable.

## Validity of Simulations

The road layout developed in this section provides a solid and repeatable basis to test the theoretical AFRC algorithm. By providing a system that is absolutely repeatable, any tests will highlight the changes that are directly caused by the control algorithm and not as a result of minor errors, omissions or approximations.

Section 1.7 provided an explanation of the “Derby Junctions” with justified the reasons for not developing this work solely around that specific area. Much of the reasoning provided is still valid to justify the choice of simulation strategy and simulation layout.

There are a number of reasons for not attempting to simulate an actual road system, some of which are related to practical limitations of available data, some to economic and licencing but the major issue is the validity and repeatability of the final results. The objective of this research is to define and prove a control concept rather than discuss the issues related to a single region. A real road system is a collection of roadways and junctions. The local problems for any specific junction are engineering problems that can be uniquely adjusted through the application of engineering practice. The solution obtained will however apply only to the specified junction at a specific time and will apply only for a period before social and economic changes cause variation in flow patterns that will necessitate further tuning. The intention of this research is to identify a new general strategy for controlling traffic at sequential active traffic control points under the very specific conditions of temporal congestion without the need for specific tuning. Consistently predicting and sustaining the level of congestion for a predetermined investigation period is not possible for a real road system as there are numerous uncorrelated events that combine at a specific temporal instance to generate a specific flow. The theoretical solution proposed by this thesis is intended to be universally applicable and not restricted to solving the problems associated with any single junction. These objectives of general strategy, universal acceptance and sustained traffic conditions preclude the use of any specific road system.

Furthermore, a valid simulation of the roadway requires real traffic data, flow rates, signal timings and, ultimately, clearly defined measurable outcomes. Some of the information may be available from a given Local Authority but complete and precise position, timing, route and destination information of individual vehicles is unlikely to be available at any point. Without complete information, any dataset that did become available would be at best an unrepeatable approximation. Results obtained from a real world simulation would be subject to calibration adjustment to account for regional driving style differences attributable to age, experience, vehicle preference and so on. A further dataset would then be required to test the outcomes. Calibration of simulation systems is a complex topic; comparing the results of simulations undertaken in this thesis to a predetermined outcome would require justification of calibration result validity and application. This would neither support nor contradict the objectives of this thesis

Practical restrictions are imposed by cost and computing power availability. The Aimsun academic licence is limited by the maximum total length of road and the maximum number of nodes (connection points between features such as two roads, a road and a junction, entry/exit from a sliproad, lane division, etc.). These features are necessary in real road networks but may easily exceed the licence restriction without substantive simulation benefit. If the road system is simplified by removing the apparently unnecessary features, the model becomes an approximation and no longer provides a true representation of the actual system. Larger simulations require greater computing power in order to complete in a reasonable period of time, the allocation of increased levels of computing power is always a contentious issue especially when the anticipated benefit is undermined by the previous arguments.

Having stated that simulation of real systems is not beneficial for this work, it is necessary to defend the suitability of the simulations that are used and justify why these are able to support the testing of the proposed algorithm. The evolution of the testing is discussed throughout this chapter. Each enhancement of the simulation creates an additional level of realism that represents the general flow and purpose of traffic without the need to define specific users. Initial tests were conducted to determine if an interaction between modified traffic control systems was possible. The simulation was then developed to create a maximum potential for interaction across multiple junctions and finally to create multidirectional tidal flow.

The overall objective of the simulation is to create a significant test of the capability of the AFRC control system using a variety of challenging flow rates and directions. The tests provide a degree of realism to show the practical use of the system but also to ensure that the test is not biased in favour of AFRC by the use of totally artificial scenarios. Features of the final simulations include:

* Multiple traffic flows co-exist at any time. Traffic does not move as a block, each vehicle will be making an individual journey or trip.
* Traffic is internal to the network. Real traffic does not simply traverse an area, some flows begin and/or end outside of the network however many vehicles begin and terminate their existence inside the network. When considering traffic flow, a vehicle is only relevant (to the extent that it only “exists”) whilst it is actively a part of the flow.
* Multiple sources and destinations. Traffic originates and terminates from various places inside the simulation. The simulation includes three distinct areas that represent residential areas. The areas imply different demographics such as a high density fully employed with access to a vehicle shared by multiple occupants of the same household; lower density fully employed with sole access to a vehicle and low density part-time, retired and semi-retired with access to a vehicle. There is also an area representing a commercial centre.
* Tidal flows. The flow rates are configured to represent a morning rush hour, an active day, an evening rush hour and a quiet night time period. The overall intention of the traffic simulation is that, over the duration of the test, all vehicles will return to their original starting location i.e. a vehicle leaving “home” will return “home” by the end of the simulation; routes and flows will be tested outbound and inbound.

A summary of the tidal flow concept is given by the following:

* Morning rush hour. Traffic flows towards the commercial area in the morning rush hour. This creates a funnel effect where the traffic density increases nearer to the commercial area. Some of the vehicles also return from the commercial area immediately towards the residential area (representing the shared use vehicles). This creates a level of traffic that is in direct opposition to the natural tidal flow, creating opposing queues and forcing signals to change with greater frequency than they would if the flow was completely unidirectional. Traffic from the low density area is minimal during the morning rush hour.
* Active Day. Traffic flow reduces during the day, there is a relatively modest flow between each area. This is intended to represent social travel as well as e.g. travel between home and shops. For the simulation, this period allows the rush hour traffic to clear while maintaining a flow to activate the traffic light systems. The AFRC concept should not provide an advantage during this period as there should be zero congestion. Conversely this period should prove that the actuated traffic signals are providing an advantage over the fully timed systems which is an important indication that the actuated model is functioning as predicted. Simultaneously, this period proves that the AFRC system does not reduce traffic flow when the system is below capacity.
* Afternoon rush hour. The majority of the traffic begins from the commercial area and then disperses to the residential areas. In this period the highest traffic density will be at the origin, congestion will reduce as traffic nears its destination. The AFRC effect will be greater at the beginning of this period, reducing as the period ends. This period shows that the traffic signals do not require any form of timing modification (tuning) in order to provide an operational advantage when the tidal flow is opposite to the original flow. This period also shows that because tuning is not necessary, the AFRC algorithm will improve traffic flow even when the flow is abnormal due to events such as nearby incidents causing unanticipated flow diversions.
* Evening / night flow. The traffic level reduces to zero over a period as a means to identify when the system is clear of all vehicles.

The main limitation of the flow concept in conjunction with the requirement to return to the originating point is that vehicles can mathematically make the journey in reverse. Numerically it is possible for a vehicle to make a return journey before the outward journey but this is not considered to be a significant problem, greater benefit is derived from considering the flow in both directions.

## Choice of Simulator

Numerous traffic simulation packages were considered at the start of this project. Vissim and Aimsun are popular commercial traffic simulation packages while Sumo [121] is an active open source project which can be used alone or with front-end packages. Commercial software frequently offers significant feature advantages while open-source packages are easier to modify at minimum cost.

Sumo, the prime open source candidate, did not support driving on the left-hand side of the road which limited its suitability for simulation in the UK and other parts of the World. The Sumo originators were contacted and confirmed at the start of this project that they were not considering changing this position. Attempts were made at working around the limitation including using mirror results and combining one-way lanes to make left-hand orientated roads however this continued to identify greater problems including road positioning during turns, priority of vehicles during turns, overtaking and lane occupancy of multi-lane roadways. This led to the decision to investigate commercial software and abandon the used of Sumo. It is understood that the decision not to include driving on the left-hand side was reversed in version 0.24.0 dated September 2015.

TSS, the original providers of Aimsun, and PTV Group who provide Vissim have extensive lists of recommendation from customers who are highly influential in transportation fields. The products offer highly graphical outputs, can be installed on a wide range of Microsoft operating systems and offer demo versions, training and support. Aimsun can also be installed in Linux. TSS and PTV offer support programmes that provide a free software licence for use in research projects. However Aimsun offered an initial 3 year licence while Vissim was an initial 1 year licence and the guarantee of continuity of use was highly influential in the decision process.

Limited experience of traffic simulation at the early stage of the project meant that there was minimal technical choice to be made between the Aimsun and Vissim with both systems having graphical outputs and capable of providing both left and right-hand drive systems. Demo versions of both systems were installed. On initial use it became clear that Aimsun was a very versatile program with a Python API that allows for control to be developed externally and applied to influence the traffic flow without needing to modify any of the other simulator functions.

Aimsun was chosen based on the licence and early use capability. The Aimsun user-group and the TSS staff in Spain and the UK have been very helpful during this project. As part of the licence agreement with Aimsun, there was a requirement not to use the licence to compare simulation systems and so no further comparisons were made.

## Simulation Host System

The following section identifies the major components of the computer system that was used for the simulation experiments.

### Hardware:

Although the actual hardware may affect the overall completion time for a simulation, it should have very minimal, if any, impact on the simulation results. However, for completeness and to allow this work to be recreated for comparison in the future, the following is a list of the major components.

Processor: AMD FX6300 6 core at 3.5GHz

Memory: 16 gigabytes, DDR3 arranged as 2 banks of 8GB

Disk Drive: 1 terabyte, sata3, physical disk, 64 megabyte cache capacity

### Software

#### Operating system:

Windows 7 64bit with service pack 1. Note: During the research period Windows 8 and Windows 10 were available as “upgrades”. For the sake of guaranteed compatibility and consistency of operation throughout the research these were declined.

#### Simulator:

Aimsun 8.0.8 (r33395 x64) with restricted capability microscopic licence intended for academic research. Initial testing was completed using an earlier release, the upgrade was necessary to resolve a function limitation. All testing was repeated with the later version to ensure consistency and further updates were declined.

#### Other software:

Python 2.7 IDE was already installed on the PC before the simulation package. Aimsun uses python2.7, a version of which is installed with the Aimsun software

AVG antivirus was installed.

## General Setup of Experiments

Each junction has an identical traffic light controller, detecting and reacting exactly the same as each other junction. This is one of the design goals for the controller and for the control algorithm; there should be no prior knowledge of the junction or flow direction. The exact controller for each experiment is shown in the “controller” section, the python code for each controller is given in the appendices.

The initial conditions of each experiment were kept identical as far as possible. In the case of this project, a major advantage to software simulation is the unrealistic ability to obtain repeatable and identical results for given inputs. This is possible because all of the initial conditions are fully defined and controllable, chaos theory does not apply; there are no truly random events. In order to simulate real life circumstances such as variation in reaction times between vehicles, the simulation package uses pseudo-random numbers generated at run time. The “random” numbers are actually a sequence of known or predictable values which is long enough to be considered non-repeating over the time period of interest. The useful sequence is made more random by selecting different seed values or starting points within the sequence. The choice of seed value can be based on many things such as the fraction of second in which the simulation starts, the position of the mouse pointer, and so on. The seed can also be set to a fixed value to ensure that the same sequence is used, guaranteeing experiment repeatability. As previously stated, with the same system inputs (demand, road network, etc.) and the same simulator seed value, the results will be identical for each simulation run. There will only be a change in outcome when there has been a change in input and the outcome change can be specifically attributed to input change if all other attributes are constant.

For the actual simulations all values are left as system defaults with the exception of the random seed being set as 10000, rerouting cycle 6 minutes and a 15 minute warm up using scenario action.

The choice of cycle time and warm up period means that simulation will be preloaded with traffic based on the route determination algorithm. The routing algorithm will have completed three times before the simulation begins (initial, 6 minutes and 12 minutes) and will complete again three minutes after the simulation starts. This was intended to alleviate any questions regarding whether the rerouting cycle would complete.

Each experiment is repeated three times with the same input parameters. The difference between each test is to enable or disable the detectors to force the controller to operate as Timed, VA or AFRC.

### Experiment Description Tables

Each experiment has a set of named traffic flow tables and an information summary table. An example summary table is shown in Figure 6.1. The table gives information to clarify the exact nature of the experiment and would be necessary to repeat an experiment.

Identifier: to clarify which experiment is being referred to, logically sequential. Not related to any simulator parameter.

Layout version: as declared in section 5.4.2.3

Aimsun Definition file: Saved version of the experiment, contains layout, flows and timing. Saved immediately before the experiment was run, provides a level of version control for the experiments.

Controller file: the Python implementation defining the signal controller in use for the experiment.

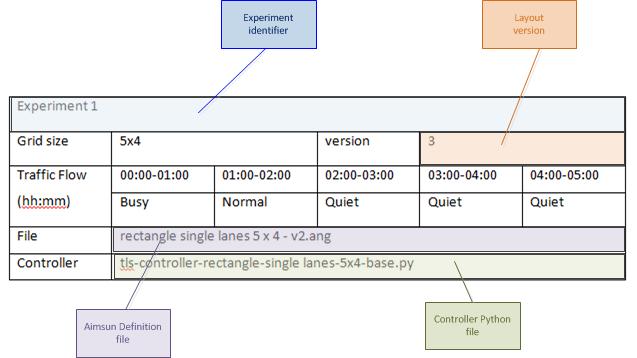


Figure .1: example description table

## Experiment 1: 5x4 Grid

Table .1: Experiment 1 Traffic Flow - Normal

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id:name | 775: South-East | 783: North-West | 1446: South-West | 1449: North-East | Total |
| 775: South-East | 0 | 1001 | 0 | 0 | 1001 |
| 783: North-West | 1000 | 0 | 0 | 0 | 1000 |
| 1446: South-West | 0 | 0 | 0 | 1000 | 1000 |
| 1449: North-East | 0 | 0 | 1000 | 0 | 1000 |
| Total | 1000 | 1001 | 1000 | 1000 | 4001 |

Note: The value of 1001 included in this experiment was initially a mistype. It was allowed to stand to confirm that very small changes are insignificant to the results. This decision also influenced later tests intended to prove that there is no pre-calculated best solution.

Table .2: Experiment 1 Traffic Flow - Busy

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id:name | 775: South-East | 783: North-West | 1446: South-West | 1449: North-East | Total |
| 775: South-East | 0 | 4000 | 0 | 0 | 4000 |
| 783: North-West | 4000 | 0 | 0 | 0 | 4000 |
| 1446: South-West | 0 | 0 | 0 | 4000 | 4000 |
| 1449: North-East | 0 | 0 | 4000 | 0 | 4000 |
| Total | 4000 | 4000 | 4000 | 4000 | 16000 |

Table .3: Experiment 1 Traffic Flow - Quiet

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id:name | 775: South-East | 783: North-West | 1446: South-West | 1449: North-East | Total |
| 775: South-East | 0 | 10 | 0 | 0 | 10 |
| 783: North-West | 10 | 0 | 0 | 0 | 10 |
| 1446: South-West | 0 | 0 | 0 | 10 | 10 |
| 1449: North-East | 0 | 0 | 10 | 0 | 10 |
| Total | 10 | 10 | 10 | 10 | 40 |

Table .4: Experiment 1 Information Summary

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Experiment 1 | | | | | |
| Grid size | 5x4 | | version | 3 | |
| Traffic Flow  (hh:mm) | 00:00-01:00 | 01:00-02:00 | 02:00-03:00 | 03:00-04:00 | 04:00-05:00 |
| Busy | Normal | Quiet | Quiet | Quiet |
| File | rectangle single lanes 5 x 4 - v2.ang | | | | |
| Controller | tls-controller-rectangle-single lanes-5x4-base.py | | | | |

This version has traffic flows across opposite corners, traffic at NE goes to SW, from NW to SE, from SE to NW and from SW to NE. Traffic is same level in all directions.

Traffic light is either fixed timing, stop line detect (VA style) or stop line and exit detect. The control code is the same for all three tests, minimum change is applied. The exit detector values are set to zero for VA, the exit and stopline detectors are set to zero for timed (runs to maximum green time)

AFRC Controller (written in python for this project) is “bang-bang” style. Each junction is read, tested and modified in numeric order (not layout order). If a blocked exit is found this direction at this light changes to red, next downstream light is forced to green from this direction regardless of current state.

Layout, flow and reaction are totally unrealistic but intended to cause maximum junction conflicts and interactions. Real traffic is very unlikely to flow equally in opposing directions and is equally unlikely to flow across a region without stopping. The “bang-bang” style refers to the way that the control will force downstream lights to change immediately regardless of the existing flow state at the junction.

### Results:

The main parameter that was being considered was Delay across the network. This is one of many parameters that can be provided by Aimsun and appeared to be the most relevant at the time of the experiment.

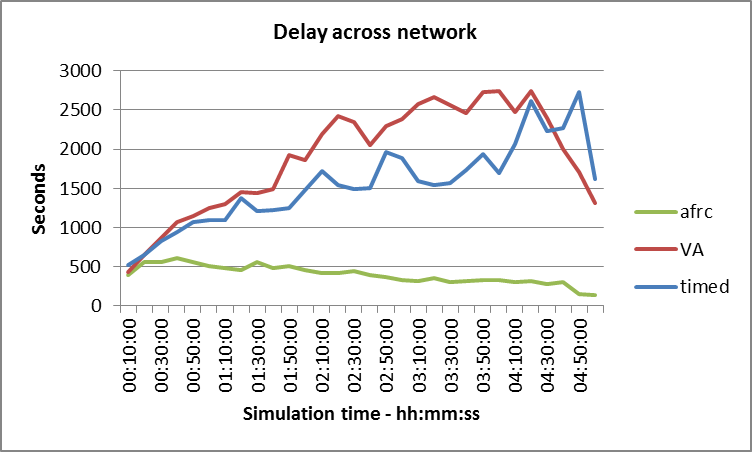


Figure .2: Experiment 1 Time Delay across Network

Delay across the network is given by

Equ( ‑)

Where *i* is the total number of vehicles inside the network at any instance and *j* is each vehicle in the range 0 to *i*.

The tabular and graphical table give the very clear impression that AFRC will provide a dramatic improvement in traffic flow over both Timed and VA control. Although impressive, this initial result was considered to be totally unrealistic due mainly to the idealised and therefore simplistic controller model.

Table .5: Experiment 1 Results - Time Delay (sec)

| time (h:m:s) | timed | VA | afrc |
| --- | --- | --- | --- |
| 00:10:00 | 516.24 | 426.51 | 393.77 |
| 00:20:00 | 649.91 | 648.78 | 559.8 |
| 00:30:00 | 821.55 | 858.61 | 554.09 |
| 00:40:00 | 942.87 | 1064.59 | 604.53 |
| 00:50:00 | 1065.81 | 1139.49 | 561.54 |
| 01:00:00 | 1091.19 | 1247.2 | 509.51 |
| 01:10:00 | 1097.32 | 1294.72 | 482.92 |
| 01:20:00 | 1371.88 | 1445.59 | 455.93 |
| 01:30:00 | 1208 | 1433.94 | 556.81 |
| 01:40:00 | 1217.47 | 1488.18 | 483.99 |
| 01:50:00 | 1242.62 | 1928.19 | 499.76 |
| 02:00:00 | 1482.33 | 1866.55 | 452.97 |
| 02:10:00 | 1719.25 | 2196.64 | 412.2 |
| 02:20:00 | 1537.5 | 2421.15 | 413.37 |
| 02:30:00 | 1492.47 | 2347.11 | 440.46 |
| 02:40:00 | 1502.23 | 2051.19 | 385.7 |
| 02:50:00 | 1965.94 | 2296.22 | 361.36 |
| 03:00:00 | 1880.36 | 2387.62 | 328.6 |
| 03:10:00 | 1597.94 | 2581.74 | 317.67 |
| 03:20:00 | 1541.48 | 2660.91 | 351.4 |
| 03:30:00 | 1567.33 | 2559.2 | 294.55 |
| 03:40:00 | 1734.22 | 2456.28 | 316.59 |
| 03:50:00 | 1934.87 | 2724.13 | 329.06 |
| 04:00:00 | 1693.6 | 2739.7 | 326.92 |
| 04:10:00 | 2066.6 | 2480.26 | 295.31 |
| 04:20:00 | 2611.34 | 2748.64 | 309.46 |
| 04:30:00 | 2230.41 | 2402.38 | 277.62 |
| 04:40:00 | 2267.53 | 2000.62 | 301.44 |
| 04:50:00 | 2731.95 | 1708.48 | 153.84 |
| 05:00:00 | 1615.24 | 1315.76 | 133.46 |

Given that the vehicle input and response is identical in each section of the experiment it was initially unclear as to why the delay increases slightly for the AFRC and peaks at approximately 40 minutes while the delay continues to rise for both VA and Timed for at least 4 hours from the start of the experiment. This is especially interesting when it is noted from the traffic profile that the flow rate of 4000 vehicles/hr should exceed the road capacity of approximately 1800 – 2500 vehicles/hr determined in section 3.3.1 for the first hour, reduce below capacity for the second hour and the reduce to a negligible flow for hours 3 – 5. As described in sect 6.5, there was also a warm up or preloading time of 15 minutes to ensure that the traffic was already inside the system when the recorded simulation began. It was concluded that there may be a restriction at the vehicle origin points where traffic was being prevented from entering the network but then as time progresses traffic is not being permitted to exit due to the traffic entering. The queues were causing spillback blocking but the action of the exit detection in the AFRC system was clearing the exit route first, ensuring minimum delay. This identifies two other significant points firstly that gating systems (delaying entry to the system, section 3.4.3) can extend the duration of congestion and secondly that the congestion / delay does not reduce in synchronisation with the time that the traffic input reduces.

An interesting aspect is that the Timed controller offers better response than the VA controller for the majority of the experiment. This initially appears to be incorrect however the previous research and theoretical control process visualised by Figure 3.13 and Table 3.13 show that a greater number of signal transitions gives rise to more total loss time and hence greater delays. The description of VA systems in section 2.3.4.2 clearly identifies that VA systems, by design, will transition more frequently than Timed systems whenever there are vehicles in the system. The improvement in response to individual vehicles at the detector in a single direction is at a cost of overall delay across the network.

### Summary

From a practical aspect the implementation is flawed in many respects, e.g. there are dependency errors in the code: the lights are tested and set in numeric sequence; if a lower numbered light forces a change at a higher numbered light, the higher numbered light may be changed again when it is processed and remove the change demanded from upstream. Similarly a higher numbered light may change to a state which causes a block to a lower numbered light.

For this experimental layout, the lights all follow the same sequence but are not all set to same timing values. The lights at the edge of the simulation have only three approach directions whereas the central lights have four. To compensate, the timing for missing approaches is set to very low values, effectively skipping the phase without having to apply a different controller at each junction.

The entry and exit points cause greatest restriction mainly because there is no effective upstream detection. The queues are allowed to develop to any length outside of the simulation which is the precise problem that this project addresses.

While these problems are numerous, the limitations should not be over emphasised. As an early stage development proof of concept model, the results are excellent; the concept works, the main limitations have been identified and further models can be developed to address the issues.

## Experiment 2: 7x6 Grid

### Setup

This was an Intermediate simulation development. As described in section 5.4.2.3 and experiment 1, there were restrictions in traffic entry for the 5x4 grid and differences in the physical layout of junctions, most having 4 approaches but some having only 3. As a result, the layout was redesigned and an additional layer of junctions provided. The new junctions have exits that have infinite routing distance and will not be selected by the simulator, no vehicles will use these exits. It was assumed that because the new junctions are not in a direct route from Origin to Destination, the amount of traffic would be light and would not have a major impact on the results. As a further part of the layout change, the sequencing of the traffic lights in the junctions was modified to a more fluid order (from the naïve sequencing of Table 2.3 to the sequencing created in Table 4.1.

It was stated that traffic is not likely to transition across the network without beginning or ending inside the system. For this reason, the Origin and Destination were relocated to a position inside the final junction rather than outside. The traffic flow is still the same level in each direction but the levels and names have been redefined for clarity; a new flow of “none” is defined in an attempt to identify the point when all traffic flow has stopped. In conjunction with this, the overall experiment time has been increased. In an apparent contradiction to an earlier statement, the “none” flow, while initially defined over an hour, can be used over any period length because the results will be identical; zero cars over 1 hour is the same flow rate as zero over 30 minutes or 3 hours. This would not be the case with e.g. the “normal” flow definition where 4000 vehicles over 1 hour is clearly not equivalent to 4000 vehicles over 3 hours.

The control code is modified to be more efficient but still works in the same way i.e. forces a next downstream change regardless of junction state. Logical disabling of detectors through a single flag is implemented to ensure identical code is used for Timed, VA and AFRC.



Figure .3: 7x6 Grid Repeated for Clarity

Table .6: Experiment 2 Information Summary

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Experiment 2 | | | | | | |
| Grid size | 7x6 | | version | 4 | | |
| Traffic Flow  (hh:mm) | 00:00-01:00 | 01:00-02:00 | 02:00-03:00 | 03:00-04:00 | 04:00-05:00 | 05:00-08:00 |
| Rushhour | Normal | Quiet | Quiet | Quiet | none |
| File | rectangle oversize 7 x 6 - v1.ang | | | | | |
| Controller | tls-controller-rectangle-oversize-7x6-v1.py | | | | | |

Table .7: Experiment 2 Traffic Flow - Normal

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id:name | 775: South-East | 783: North-West | 1446: South-West | 1449: North-East | Total |
| 775: South-East | 0 | 1000 | 0 | 0 | 1000 |
| 783: North-West | 1000 | 0 | 0 | 0 | 1000 |
| 1446: South-West | 0 | 0 | 0 | 1000 | 1000 |
| 1449: North-East | 0 | 0 | 1000 | 0 | 1000 |
| Total | 1000 | 1000 | 1000 | 1000 | 4000 |

Table .8: Experiment 2 Traffic Flow – Rushhour

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id:name | 775: South-East | 783: North-West | 1446: South-West | 1449: North-East | Total |
| 775: South-East | 0 | 4000 | 0 | 0 | 4000 |
| 783: North-West | 4000 | 0 | 0 | 0 | 4000 |
| 1446: South-West | 0 | 0 | 0 | 4000 | 4000 |
| 1449: North-East | 0 | 0 | 4000 | 0 | 4000 |
| Total | 4000 | 4000 | 4000 | 4000 | 16000 |

Table .9: Experiment 2 Traffic Flow - Quiet

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id:name | 775: South-East | 783: North-West | 1446: South-West | 1449: North-East | Total |
| 775: South-East | 0 | 10 | 0 | 0 | 10 |
| 783: North-West | 10 | 0 | 0 | 0 | 10 |
| 1446: South-West | 0 | 0 | 0 | 10 | 10 |
| 1449: North-East | 0 | 0 | 10 | 0 | 10 |
| Total | 10 | 10 | 10 | 10 | 40 |

Table .10: Experiment 2 Traffic Flow - None

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id:name | 775: South-East | 783: North-West | 1446: South-West | 1449: North-East | Total |
| 775: South-East | 0 | 0 | 0 | 0 | 0 |
| 783: North-West | 0 | 0 | 0 | 0 | 0 |
| 1446: South-West | 0 | 0 | 0 | 0 | 0 |
| 1449: North-East | 0 | 0 | 0 | 0 | 0 |
| Total | 0 | 0 | 0 | 0 | 0 |

### Results

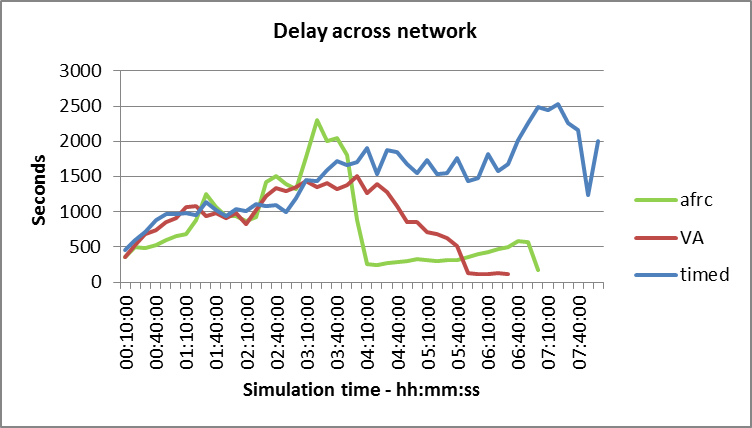


Figure .4: Experiment 2 Time Delay across Network

The results of this test appear to show that the AFRC system is not providing the positive results indicated in the previous experiment. Neither system, AFRC or VA, displaying a clear advantage.

From the graphical results no system appears to have a significantly better response over the first 2.5 hours however it is important to observe the tabular data to understand the implications. It is comparatively easy to discount many of the points on the graph as unexceptional, especially before 03:00 however the tabular data for e.g. 02:30:00 shows delays of 1080 sec, 1227 sec and 1415 sec for Timed, VA and AFRC respectively; approximately 3 minutes difference between each system.

The timed results are not totally unexpected, reasonably good flow with the peak traffic volume but very poor as the volume reduces. Vehicles are being held for long periods at each signal on the route, with no synchronisation of signals and no route priority, vehicles may be held at every junction along the route. It was a surprising feature that the traffic does not appear to completely clear the network even after 3 hours without additional traffic.

Table .11: Experiment 2 Results - Time Delay (sec)

| time(h:m:s) | timed | VA | afrc |
| --- | --- | --- | --- |
| 00:10:00 | 462.21 | 363.35 | 361.55 |
| 00:20:00 | 593.7 | 528.73 | 493.01 |
| 00:30:00 | 706.97 | 683.02 | 488.47 |
| 00:40:00 | 887.93 | 734.34 | 523.27 |
| 00:50:00 | 967.23 | 856.11 | 594.07 |
| 01:00:00 | 961.83 | 909 | 656.87 |
| 01:10:00 | 981.52 | 1060.17 | 688.15 |
| 01:20:00 | 945.97 | 1080 | 884.37 |
| 01:30:00 | 1143.85 | 941.55 | 1257.18 |
| 01:40:00 | 1020.34 | 986.25 | 1070.03 |
| 01:50:00 | 945.15 | 910.22 | 936.53 |
| 02:00:00 | 1042.61 | 975.63 | 939.74 |
| 02:10:00 | 1012.68 | 824.88 | 873.7 |
| 02:20:00 | 1110.97 | 1017.14 | 918.05 |
| 02:30:00 | 1080.08 | 1227.06 | 1414.62 |
| 02:40:00 | 1089.82 | 1332.46 | 1501.28 |
| 02:50:00 | 997.05 | 1295.81 | 1392.15 |
| 03:00:00 | 1193.68 | 1349.11 | 1325.8 |
| 03:10:00 | 1449.09 | 1430.72 | 1782.43 |
| 03:20:00 | 1434.34 | 1351.89 | 2297.61 |
| 03:30:00 | 1595.47 | 1406.86 | 2004.77 |
| 03:40:00 | 1726.51 | 1326.78 | 2053.35 |
| 03:50:00 | 1664.59 | 1373.12 | 1810.29 |
| 04:00:00 | 1711.74 | 1503.14 | 883.11 |
| 04:10:00 | 1908.18 | 1268.38 | 251.35 |
| 04:20:00 | 1533 | 1388.68 | 247.07 |
| 04:30:00 | 1869.43 | 1282.17 | 275.01 |
| 04:40:00 | 1842.97 | 1077.32 | 290.53 |
| 04:50:00 | 1682.53 | 853.68 | 302.94 |
| 05:00:00 | 1548.68 | 855.31 | 322.5 |
| 05:10:00 | 1737.22 | 707.47 | 311.33 |
| 05:20:00 | 1537.2 | 686.79 | 299.63 |
| 05:30:00 | 1547.72 | 627.8 | 307.41 |
| 05:40:00 | 1761.46 | 512.05 | 313.83 |
| 05:50:00 | 1433.35 | 128.04 | 360.57 |
| 06:00:00 | 1485.43 | 114.05 | 395.83 |
| 06:10:00 | 1817.86 | 109.5 | 429.56 |
| 06:20:00 | 1575.77 | 123.67 | 466.39 |
| 06:30:00 | 1680.3 | 118.15 | 495.27 |
| 06:40:00 | 2017.8 |  | 579.22 |
| 06:50:00 | 2256.8 |  | 565.89 |
| 07:00:00 | 2486.43 |  | 165.14 |
| 07:10:00 | 2448.99 |  |  |
| 07:20:00 | 2531.73 |  |  |
| 07:30:00 | 2262.61 |  |  |
| 07:40:00 | 2157.93 |  |  |
| 07:50:00 | 1241.94 |  |  |
| 08:00:00 | 2009.84 |  |  |

There is a significant increase in delay for the in the AFRC system between 03:10:00 and 03:40:00 followed by the very significant drop before 04:10:00.

The delay for VA reduces progressively from approximately 04:00:00 while the AFRC delay increases. The peak delay for the timed system is near to the end of the experiment, there is a local peak in the delay for the AFRC system at a similar time.

The delay for the VA system is lower than the AFRC after 05:40:00 where it stays for the remainder of the experiment. The delay for VA reaches zero 30 minutes sooner than for AFRC.

### Summary

Similarities were expected between the initial response of VA and AFRC due to the exit detection being unnecessary / not beneficial unless the network is congested however the increasing level of delay implies that there network is congested.

The reasons for the increasing delay in the AFRC system between 03:10:00 and 03:40:00 followed by the very significant drop to 04:10:00 were not easily explained. From careful examination of the data and an amount of experimentation within the controller code it was determined that although the light sequence was modified through a layout change, the action of AFRC also changes the sequence when a blocked exit is identified. The controller attempts to find a next suitable exit however this may not be optimal.

The suitability and definition of delay across the network was considered. The definition in Equ( 6‑1) shows that the figure obtained is a summation of the delays encountered by all of the vehicles within the network at the instance of measurement. This is a compound result, it cannot be determined from this data alone whether a single vehicle is experiencing one large delay or whether a large number of vehicles are each experiencing a small delay. This understanding may clarify the results. For example, the Timed result could be a very small number of vehicles experiencing excessive delays at each signal, the delays for each vehicle would be compound and the overall result compounded again. Similarly the sudden drop in the AFRC result could be attributed to a small number of vehicles exiting the system that had previously experienced large delays.

Overall, the traffic flow is unrealistic but some of the layout and timing issues were resolved. Entry and exit points continued to cause greatest restriction; they were not resolved by this solution.

## Experiment 3: 7x6 realflow grid

### Setup

The controller was modified to read the detectors at the end of each simulation cycle. This change in methodology ensured that the detection information is available before the start of the subsequent cycle for every cycle except the first. Having detection data available for all signals before it is required allows the local junction to determine a signal response including the upstream exit detection; previous versions forced an immediate change in the downstream signal in direct response to the exit detector regardless of the current stage or timing of the current stage of the downstream signal. The apparent problem of zero detection data at the start of the simulation cycle has no impact. Detection data will be zero for the first cycle during the warm up period; warm up cycles are conducted exactly as other cycles, including signal action, with the exception that the outcomes are not included in the experiment results. The effect of missing data / zero detection for the first cycle will be that signals would not change before the maximum green time which is many cycles later. Even if the detection data was actually read, it would still be zero before cycle 1 as there would be no vehicles inside the system and so no detections would be possible. The situation is resolved when detection data becomes available at the end of cycle 1 for inclusion in cycle 2, when the data will almost certainly remain at zero.

The traffic input matrix is modified significantly. The changes are intended to show how the AFRC system is durable in the ability to manage variation in junction approach flow volume and direction without making changes to sequencing or base timing. Successful testing should indicate the ability of AFRC to adapt to unforeseen and immediate changes such as those caused by accidents, incidents and roadworks.

Simultaneously and in recognition of real world environments, the scenario changes are indicative of bulk traffic flow changes through the duration of a normal day. The modified traffic demand consists of the same overall traffic volume but approaching the network from different directions. The traffic flows are intended to represent traffic from three residential areas towards a commercial / mass employment area as discussed in 5.4.2.4. The size and demography of the areas is different and reflected by the traffic flow. Flow rates and destinations are not equal in every direction and change during simulation period. One destination attracts more traffic than all others, intended to simulate a mass employment destination. Other destinations intended to simulate common items such as school run, daily shopping, etc. by both timing and destination.

Traffic source and destination is now part way along street, not at junction – no longer directly determined by individual traffic light.

The simulation outputs status information to a .csv file at the end of each cycle including detector states and light transitions. The data files are large, each csv line has 27 values and a new line is generated for each junction every 0.8 simulation seconds which is equivalent to a rate of 3150 lines per simulation minute or 189,000 lines per simulation hour. Current versions of MS Excel (2013-2016) support slightly more than 1 million lines which means a simulation of 5.5 hours could be opened, manipulated and examined albeit very slowly. The data files were useful to identify the transitions of any particular junction, which was necessary for later developments. A small sample of the data is presented as an example in the appendix but the files are too large to be included in their entirety.

Flow rates and destinations are not equal and change during sim period. One destination attracts more traffic than all others, intended to simulate a mass employment destination. Other destinations intended to simulate common items such as school run, daily shopping, etc. by both timing and destination.

There is a change to the layout in response to the incoming and outgoing flow restrictions; the traffic Origin and Destinations have been relocated to a position which is part way along a route section rather than at a junction. This means that the input and output are no longer determined primarily by an individual traffic signal. Traffic is permitted to routes that are completely outside of the direct paths, in the earlier version traffic could only exceed the direct route in East and West directions.

To aid with post simulation understanding, the controller simulation writes out status information to a file in comma separated variable (.csv) format at the end of each simulation cycle including detector states and signal transitions. Each line of the csv file consists of 27 values. A new line is generated for each junction every 0.8 simulation seconds which is equivalent to a rate of 3150 lines per simulation minute or 189,000 lines per simulation hour. Current versions of MS Excel (2013-2016) support only slightly more than 1 million lines which means a simulation of approximately 5.25 hours could be opened, manipulated and examined albeit, relatively speaking, very slowly. The data files were useful to identify the transitions of any particular junction, which was necessary to identify anomalies and for later developments. A small sample of the data is presented as an example in the appendix but the files have little information value without specific context and are too large to be included in their entirety.

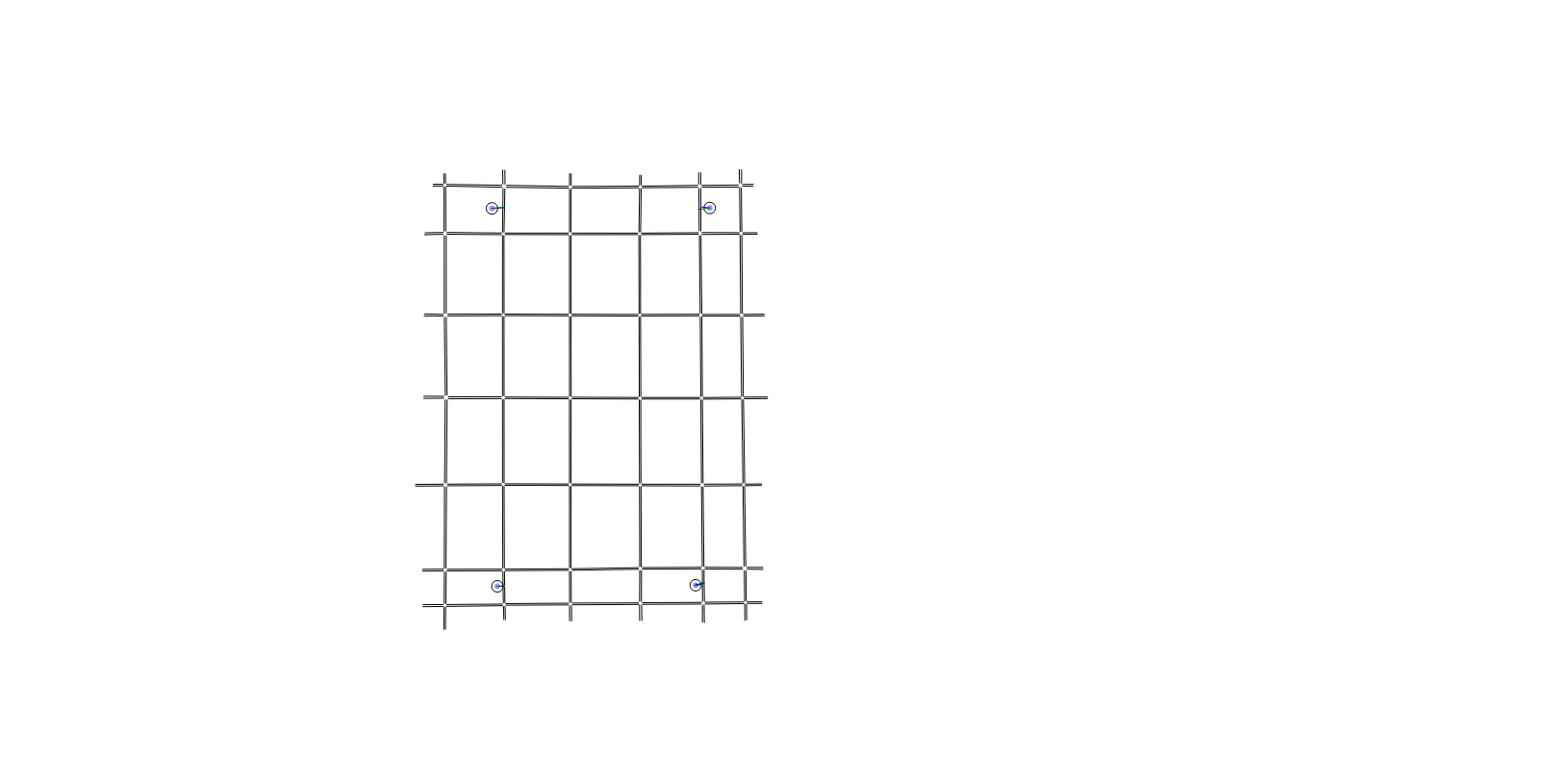


Figure .5: 7x6 Realfow Grid

Table .12: Experiment 3 Information Summary

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Experiment 3 | | | | | |
| Grid size | 7x6 | | version | 5 | |
| Traffic Flow  (hh:mm) | 00:00-01:00 | 01:00-02:00 | 02:00-03:00 | 03:00-03:30 | 03:30-4:30 |
| Rushhour | Normal | Quiet | Quiet | None |
| File | rectangle oversize 7 x 6 -realflow - v2.ang | | | | |
| Controller | tls-controller-rectangle-oversize-7x6-v5.py | | | | |

Table .13: Experiment 3 Traffic Flow - Normal

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id:name | 775: South-East | 783: North-West | 1446: South-West | 1449: North-East | Total |
| 775: South-East | 0 | 400 | 400 | 400 | 1200 |
| 783: North-West | 400 | 0 | 0 | 0 | 400 |
| 1446: South-West | 400 | 0 | 0 | 0 | 400 |
| 1449: North-East | 400 | 0 | 0 | 0 | 400 |
| Total | 1200 | 400 | 400 | 400 | 2400 |

Table .14: Experiment 3 Traffic Flow – Rushhour

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id:name | 775: South-East | 783: North-West | 1446: South-West | 1449: North-East | Total |
| 775: South-East | 0 | 100 | 5 | 75 | 180 |
| 783: North-West | 2000 | 0 | 0 | 0 | 2000 |
| 1446: South-West | 1000 | 0 | 0 | 0 | 1000 |
| 1449: North-East | 2000 | 0 | 0 | 0 | 2000 |
| Total | 5000 | 100 | 5 | 75 | 5180 |

Table .15: Experiment 3 Traffic Flow - Quiet

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id:name | 775: South-East | 783: North-West | 1446: South-West | 1449: North-East | Total |
| 775: South-East | 0 | 50 | 50 | 50 | 150 |
| 783: North-West | 50 | 0 | 100 | 100 | 250 |
| 1446: South-West | 50 | 100 | 0 | 80 | 230 |
| 1449: North-East | 50 | 100 | 80 | 0 | 230 |
| Total | 150 | 250 | 230 | 230 | 860 |

Table .16: Experiment 3 Traffic Flow - None

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id:name | 775: South-East | 783: North-West | 1446: South-West | 1449: North-East | Total |
| 775: South-East | 0 | 0 | 0 | 0 | 0 |
| 783: North-West | 0 | 0 | 0 | 0 | 0 |
| 1446: South-West | 0 | 0 | 0 | 0 | 0 |
| 1449: North-East | 0 | 0 | 0 | 0 | 0 |
| Total | 0 | 0 | 0 | 0 | 0 |

### Results

First, considering and comparing the Delay as with the previous experiments. The result is as anticipated, VA and AFRC show better results than Timed with traffic building however Timed is better than VA in at certain times which is assumed to be when there is a high level of congestion. The function of AFRC and VA are identical until there is significant congestion at which point the AFRC algorithm appears to be superior with lower overall delays and faster traffic clearance time.

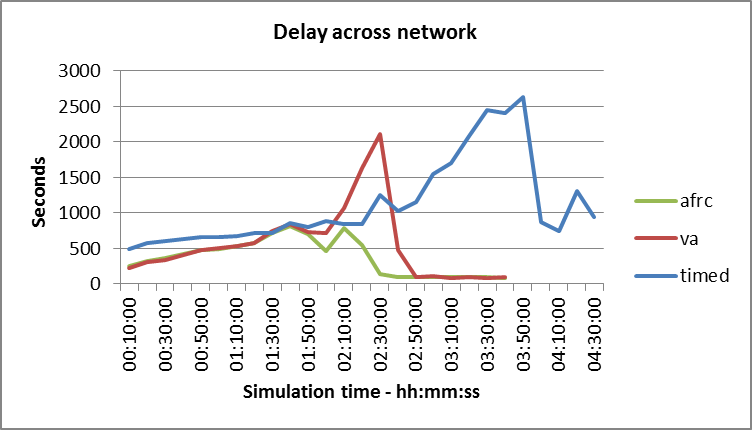


Figure .6: Experiment 3 Time Delay across Network

The results appear to be satisfactory however the previous experiment identified that the same general type of result can be achieved in several different ways. For example, it is possible to show a reduction in the network delay if the number of vehicles is reduced; the reduction could be achieved by gating or preventing vehicles from entering the network. Within the experiments there is no actual intention to prevent traffic from entering the network; in point of fact, the Origin / Destination were relocated several times specifically to reduce or avoid this problem. However, having recognised the impact of gating and the need for deeper understanding of the results, further tabular data was collected and graphically analysed. In order to show the potential for gating, the flow of vehicles into the network was examined, the results are shown in and . Higher values in this data are generally better as they indicate a less restrictive inflow.

On first inspection these results are not unexpected. There is a high influx of traffic during the first hour due the “rushhour” traffic flow definition. For VA and AFRC, the influx then tapers as the traffic reduces to the “normal flow”. The level reduces almost immediately for the duration of the “quiet” definition at 2:00:00 and all three systems reduce rapidly to zero when the “none” definition begins at 3:30:00.

Table .17: Experiment 3 Time Delay Across Network (sec)

| Time (h:m:s) | Timed | va | afrc |
| --- | --- | --- | --- |
| 00:10:00 | 485.42 | 221.84 | 253.75 |
| 00:20:00 | 578.79 | 302.51 | 320.55 |
| 00:30:00 | 600.48 | 338.86 | 361.06 |
| 00:40:00 | 632.68 | 407.15 | 412.15 |
| 00:50:00 | 662.6 | 475.24 | 470.72 |
| 01:00:00 | 657.36 | 508.35 | 490.16 |
| 01:10:00 | 678.4 | 535.08 | 524.05 |
| 01:20:00 | 708.92 | 568.93 | 579.08 |
| 01:30:00 | 715.06 | 748.84 | 716.28 |
| 01:40:00 | 860.18 | 843.73 | 813.86 |
| 01:50:00 | 794.43 | 727.93 | 693.21 |
| 02:00:00 | 879.62 | 709.7 | 454.32 |
| 02:10:00 | 847.1 | 1073.47 | 786.79 |
| 02:20:00 | 847.17 | 1636.62 | 546.59 |
| 02:30:00 | 1251.41 | 2112.03 | 133.17 |
| 02:40:00 | 1019.93 | 474.01 | 93.33 |
| 02:50:00 | 1146.88 | 89.02 | 90.69 |
| 03:00:00 | 1545.37 | 106.39 | 91.69 |
| 03:10:00 | 1695.54 | 82.53 | 94.24 |
| 03:20:00 | 2075.83 | 88.25 | 93.11 |
| 03:30:00 | 2453.93 | 81.35 | 89.6 |
| 03:40:00 | 2404.79 | 92.63 | 80.55 |
| 03:50:00 | 2627.09 |  |  |
| 04:00:00 | 872.12 |  |  |
| 04:10:00 | 735.74 |  |  |
| 04:20:00 | 1312.64 |  |  |
| 04:30:00 | 939.85 |  |  |

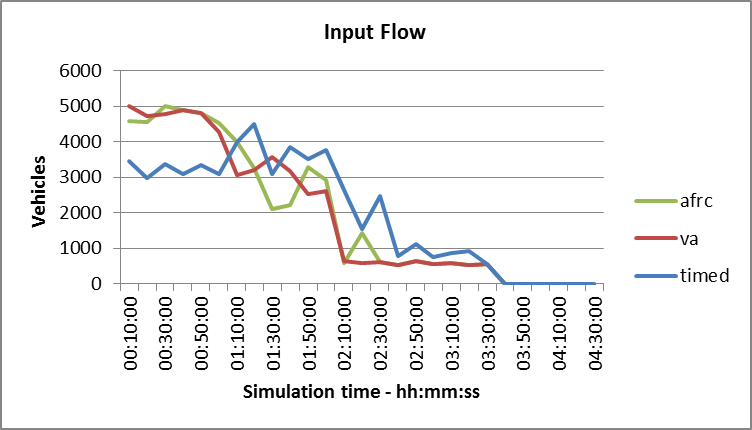


Figure .7: Experiment 3 Traffic Input Flow

The timed version has a much lower input rate than VA or AFRC during the initial period. The implication of this is that although Timed systems should offer a greater throughput due to less lost time through fewer signal transitions, they are not dynamic in response to flow changes and will create longer average delays due to the extended Red signal time. This is discussed in section , where the observations are made that the average waiting time is 0.5 x red time and that the benefit of extended green time is only useful if there are vehicles waiting to take advantage of it, i.e. they are consistently queued in advance. The differences in input flow between VA / AFRC and Timed give a strong indication that traffic must be delayed at the input for Timed as an absolute minimum.

The second observation is that the input rate increases for the Timed system when the rate for VA/AFRC is falling. The most obvious reason for this would be that the incoming traffic is becoming more consistent in the arrival rate due to buffering. The rate for VA / AFRC is reducing simply because there are fewer vehicles in the flow definition.

Table .18: Experiment 3 Traffic Input Flow (vehicles)

| Time (h:m:s) | Timed | va | afrc |
| --- | --- | --- | --- |
| 00:10:00 | 3462 | 5016 | 4596 |
| 00:20:00 | 2976 | 4716 | 4548 |
| 00:30:00 | 3360 | 4770 | 5010 |
| 00:40:00 | 3102 | 4884 | 4884 |
| 00:50:00 | 3342 | 4806 | 4818 |
| 01:00:00 | 3096 | 4290 | 4536 |
| 01:10:00 | 3990 | 3054 | 3990 |
| 01:20:00 | 4506 | 3198 | 3264 |
| 01:30:00 | 3096 | 3570 | 2106 |
| 01:40:00 | 3846 | 3186 | 2208 |
| 01:50:00 | 3504 | 2520 | 3288 |
| 02:00:00 | 3756 | 2622 | 2922 |
| 02:10:00 | 2634 | 648 | 594 |
| 02:20:00 | 1554 | 576 | 1416 |
| 02:30:00 | 2460 | 600 | 600 |
| 02:40:00 | 774 | 522 | 522 |
| 02:50:00 | 1128 | 630 | 630 |
| 03:00:00 | 762 | 552 | 552 |
| 03:10:00 | 864 | 588 | 588 |
| 03:20:00 | 930 | 534 | 534 |
| 03:30:00 | 564 | 564 | 564 |
| 03:40:00 | 0 | 0 | 0 |
| 03:50:00 | 0 | 0 | 0 |
| 04:00:00 | 0 | 0 | 0 |
| 04:10:00 | 0 | 0 | 0 |
| 04:20:00 | 0 | 0 | 0 |
| 04:30:00 | 0 | 0 | 0 |

The peaks in the AFRC correspond to instances where the input flow fall to lower levels, most notable at 02:30:00, indicating a problem with vehicle delivery and again suggesting queueing outside of the simulation.

The observations and the relative difference between the anticipated flow rate and the actual flow rate for all of the systems indicate that being held in queue input buffers that are not observable from the data already collected. This was confirmed with the “waiting to enter information shown in Figure 6.8 and .

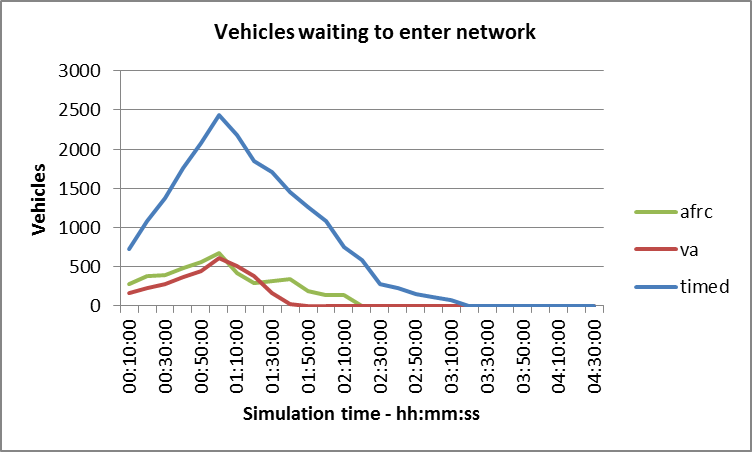


Figure .8: Experiment 3 Vehicles Waiting to Enter

Table .19: Experiment 3 Vehicles Waiting to Enter

| Time (h:m:s) | timed | va | afrc |
| --- | --- | --- | --- |
| 00:10:00 | 724 | 155 | 279 |
| 00:20:00 | 1082 | 223 | 375 |
| 00:30:00 | 1374 | 280 | 392 |
| 00:40:00 | 1757 | 366 | 478 |
| 00:50:00 | 2079 | 444 | 554 |
| 01:00:00 | 2439 | 605 | 674 |
| 01:10:00 | 2184 | 506 | 419 |
| 01:20:00 | 1843 | 383 | 285 |
| 01:30:00 | 1702 | 163 | 309 |
| 01:40:00 | 1454 | 25 | 334 |
| 01:50:00 | 1265 | 0 | 181 |
| 02:00:00 | 1076 | 0 | 131 |
| 02:10:00 | 745 | 0 | 140 |
| 02:20:00 | 582 | 0 | 0 |
| 02:30:00 | 272 | 0 | 0 |
| 02:40:00 | 230 | 0 | 0 |
| 02:50:00 | 147 | 0 | 0 |
| 03:00:00 | 112 | 0 | 0 |
| 03:10:0066 | 0 | 0 | 0 |
| 03:20:00 | 0 | 0 | 0 |
| 03:30:00 | 0 | 0 | 0 |
| 03:40:00 | 0 | 0 | 0 |
| 03:50:00 | 0 | 0 | 0 |
| 04:00:00 | 0 | 0 | 0 |
| 04:10:00 | 0 | 0 | 0 |
| 04:20:00 | 0 | 0 | 0 |
| 04:30:00 | 0 | 0 | 0 |

By comparing results it becomes clear that the timed system is performing better than the VA system because the number of vehicles in the network is lower. While fully tuned timed systems should pass more traffic and therefore offer improved traffic flow, this system is completely untuned with many sequential stops. The outcome should be that the timed system restricts the traffic flow to a greater extent than the VA system

### Summary

The experiment results clearly support the expectations of improved traffic clearance when using an exit detection system. The delay graph taken in isolation implies that VA and AFRC systems have only marginal improvement over Timed signal systems. Earlier discussion considered this as a possibility but the results of the previous experiments did not support this. The “waiting to enter” data showed that the number of vehicles in the system was much lower for the Timed system and therefore the Delay across the network data could be misleading.

Delay across the network is not sufficient to understand the relative performance of the different control systems, additional parameters are required to understand the impact that this network has on the surrounding area. Input flow is a parameter that can indicate if the experiments are managing the same levels of traffic and identifying whether the network is having an impact on the external region. The difference between input flows when the same level of traffic is being offered can be seen from “waiting to enter” data. Waiting to enter data provides a measure of the impact of traffic outside of the test region.

In examining the difference between the input flow of each system and the unusual delay characteristics, a fundamental design problem with the traffic entering the network was identified. The traffic capacity for the single carriageway was established in section to be in the range 1800 – 2500 vehicles/hr (Aimsun uses 1800 in the configuration), however the layout allows traffic to enter in either direction and so the peak input capacity for the simulation should be 3600 veh/hr. The maximum input rate from any single entrance, according to the traffic profile, is 2000 vehicles/hr and so the input queue should be much lower than the results have shown even allowing for the action of the signalling systems. The Origin / Destination centroid (the point where traffic can enter or leave) had been placed in the centre of a roadway to achieve this but was unintentionally declared as uni-directional limiting the input to 1800veh/hr. Moreover, traffic joining the network could only initially travel in one direction which meant that it was being totally controlled by the first signal in the roadway, as discussed in section signal transition losses would reduce the input capacity to significantly less than 1800 vehicles/hr with the approximate 50% duty cycle of the Timed system reducing to less than 900 veh/hr. For similar reasons, traffic could only leave the system when travelling in one specific direction and that would also be restricted by the approach signal.

To prove the AFRC capability the level of congestion has to be maximised. Without changing the layout to be artificially congested, the best method to increase the congestion is to increase the number of vehicles inside the network either by adding more Origin points or increasing the input flow capacity. The error was corrected in all subsequent tests.

The “waiting to enter” data showed an anomaly in the flow of vehicles into the network between AFRC and VA systems. Further testing identified that due to a programming error, the controller did not update the signal change requests from all directions. The problem occurred because, although the entire code is common for the controllers, the error was in a section that was only active with AFRC. This coding error was corrected for subsequent testing.

## Experiments 4-9: 7x6 realflow grid

### Setup

This group of tests is almost identical to Experiment 3 but with corrections implemented for the problems that were identified in the evaluation of the results from experiment 3.

Conducting the first of these experiments shows that the problems have been resolved. Experiment 3 was intended to show the adaptability of AFRC to traffic approaching at different rates and from different directions however minor flaws in layout design and coding did not identify the major improvements that were anticipated.

Section showed the effect of different cycle times on traffic flows. In the experiments 1 -3, the minimum and maximum green time were fixed at predefined levels. The levels were considered reasonable but were not proven. The signal timing implemented was:

Min\_green: 8 seconds. This is the time that a green signal must stay active for once it is shown. The current direction green time can expire any time after the min\_green time if no vehicles are passing the stop line detector i.e. there is no traffic flow. The value ensures that a minimum number of vehicles can enter the junction during the green sequence. This also avoids indecision and reduces the likelihood of civil disobedience (entering the junction after the signal changes to red). Does not apply to Timed systems.

Max\_green: 50 seconds. In the VA and AFRC implementation, this is the maximum time that a green will be shown in a given direction if traffic is waiting in a conflicting direction. The max\_green time is effective regardless of current direction flow. Not applicable to Timed systems.

Abs\_max\_green: 70 seconds. This is the maximum time that a green will be shown in a given direction. This value prevents an infinite red time being shown in a direction in the event of a detector failure in VA or AFRC systems. This is the duration of green time in the Timed system.

Experiments 4-8 were undertaken with various values of min\_green, max\_green and abs\_max\_green in order to identify if incidental harmonic values have been used. The ultimate objective is to show that the new algorithm is independent of the specific values; the improvements over comparative systems should remain generally consistent even though the levels change.

The final experiment in the set, experiment 9, is an average of 10 simulations using the same data but different seed values to allow for different human responses.

The experiments all use the same layout, traffic demand and controller definition, only the green time timing constants are changed.

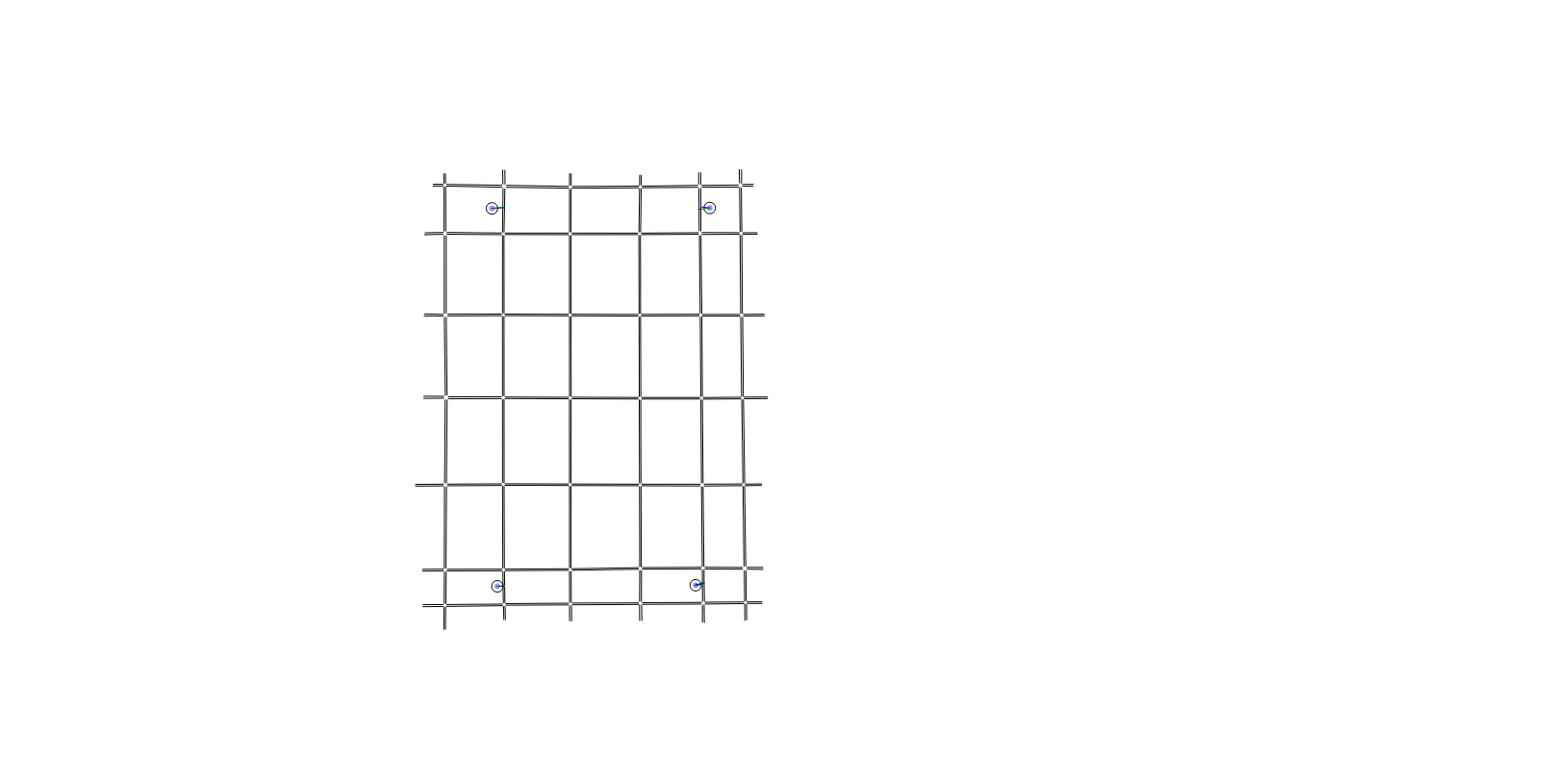


Figure .9: 7x6 Realflow Grid - Experiments 4 - 9

Table .20: Experiment 4-9 Information Summary

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Experiment 4 | | | | | |
| Grid size | 7x6 | | Version | 5 | |
| Traffic Flow  (hh:mm) | 00:00-01:00 | 01:00-02:00 | 02:00-03:00 | 03:00-03:30 | 03:30-4:30 |
| Rushhour | Normal | Quiet | Quiet | None |
| File | rectangle oversize 7 x 6 -realflow – v2.ang | | | | |
| Controller | tls-controller-rectangle-oversize-7x6-v7-working.py | | | | |

Table .21: Experiment 4-9 Traffic Flow - Normal

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id:name | 775: South-East | 783: North-West | 1446: South-West | 1449: North-East | Total |
| 775: South-East | 0 | 400 | 400 | 400 | 1200 |
| 783: North-West | 400 | 0 | 0 | 0 | 400 |
| 1446: South-West | 400 | 0 | 0 | 0 | 400 |
| 1449: North-East | 400 | 0 | 0 | 0 | 400 |
| Total | 1200 | 400 | 400 | 400 | 2400 |

Table .22: Experiment 4-9 Traffic Flow – Rushhour

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id:name | 775: South-East | 783: North-West | 1446: South-West | 1449: North-East | Total |
| 775: South-East | 0 | 100 | 5 | 75 | 180 |
| 783: North-West | 2000 | 0 | 0 | 0 | 2000 |
| 1446: South-West | 1000 | 0 | 0 | 0 | 1000 |
| 1449: North-East | 2000 | 0 | 0 | 0 | 2000 |
| Total | 5000 | 100 | 5 | 75 | 5180 |

Table .23: Experiment 4-9 Traffic Flow - Quiet

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id:name | 775: South-East | 783: North-West | 1446: South-West | 1449: North-East | Total |
| 775: South-East | 0 | 50 | 50 | 50 | 150 |
| 783: North-West | 50 | 0 | 100 | 100 | 250 |
| 1446: South-West | 50 | 100 | 0 | 80 | 230 |
| 1449: North-East | 50 | 100 | 80 | 0 | 230 |
| Total | 150 | 250 | 230 | 230 | 860 |

Table .24: Experiment 4-9 Traffic Flow - None

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id:name | 775: South-East | 783: North-West | 1446: South-West | 1449: North-East | Total |
| 775: South-East | 0 | 0 | 0 | 0 | 0 |
| 783: North-West | 0 | 0 | 0 | 0 | 0 |
| 1446: South-West | 0 | 0 | 0 | 0 | 0 |
| 1449: North-East | 0 | 0 | 0 | 0 | 0 |
| Total | 0 | 0 | 0 | 0 | 0 |

### Experiment 4 Summary:

This experiment acted as the baseline experiment for the set; it used the same initial timing values as the previous experiments. The values of interest were minimum green time, maximum green time and absolute maximum green time. These were normally set at 8, 50 and 70 seconds respectively.

Tabular outputs of delay time, stop time, travel time, number of vehicles inside the network, number of vehicles waiting to enter and total travel distance were obtained for each part of the experiment. This is additional data to the earlier experiments and is intended to simplify observation of values for e.g. vehicles inside the network which was previously being estimated / interpreted from the traffic profile and “waiting to enter” values. As previously stated, the time values obtained refer to all vehicles in the network rather than any individual vehicle, the “vehicles inside” data adds a layer of interpretation to the results obtained.

The results of the experiment are shown graphically in Figure 6.10 to Figure 6.15. The conclusions and results interpretation is given in section .

### Experiment 4 Results

Minimum green = 8 sec, Maximum green = 50 sec, Absolute maximum green=70 sec

|  |  |  |
| --- | --- | --- |
| Figure .: Experiment 4 Delay across Network | Figure .: Experiment 4 Stop Time | Figure .: Experiment 4 Travel Time |
| Figure .: Experiment 4 Vehicles Inside Network | Figure .: Experiment 4 Vehicles Waiting to Enter | Figure .: Experiment 4 Travel Distance |

### Experiment 4 interpretation and conclusions

Experiment 4 is very similar to experiment 3 but with errors corrected. The initial results from “Delay across network” () however show a distinctly different result. The reason for this can be determined from the profile of the “Vehicles waiting to enter” graph (Figure 6.14). Resolving the problem of unidirectional traffic entry in experiment 3 has reduced the input queue size and moved the peak to an earlier time.

A brief comparison of Figure 6.10: Experiment 4 Delay across Network, Figure 6.14: Experiment 4 Vehicles Waiting to Enter and Figure 6.13: Experiment 4 Vehicles Inside Network shows that delays are lower with AFRC even though more vehicles are able to enter (small queueing). The low numbers of ”Vehicles inside the network” were previously shown to artificially imply better performance however in this case fewer vehicles inside identifies that vehicles are exiting the network at a greater rate which means that their journeys must be satisfied faster, further indicating the benefit of AFRC.

Another conclusion from this is that care must be taken to consider the results in context.

An obvious observation is the similarity between delay time, stop time and travel time. The data used to create the figures is included in the appendices and shows that there are differences. The three sets of data are closely related, it is logical to assume that the delay will be related to the length of time that vehicles are stationary and that in turn, this influences the travel time. Further comparison shows that there is a constant offset of around 107 seconds/mile between the reported travel time and reported delay time. The implication of this value is that, without delays, the transit speed should be 107 seconds per mile.

Equ( ‑)

There are distinct step changes in the graphs. The primary driver for the changes is most obviously the traffic input. Figure 6.16 shows the traffic input matrix superimposed over the Delay across network graph (Figure 6.10). It is important to note that the traffic volume is the volume offered and not necessarily accepted by the network, the difference is indicated by the waiting to enter graph (Figure 6.14: Experiment 4 Vehicles Waiting to Enter.

shows that there is a step delay between the change in traffic input and the network response, e.g. the input volume reduces at 01:00 but the delay for VA continues to increase until 02:30.

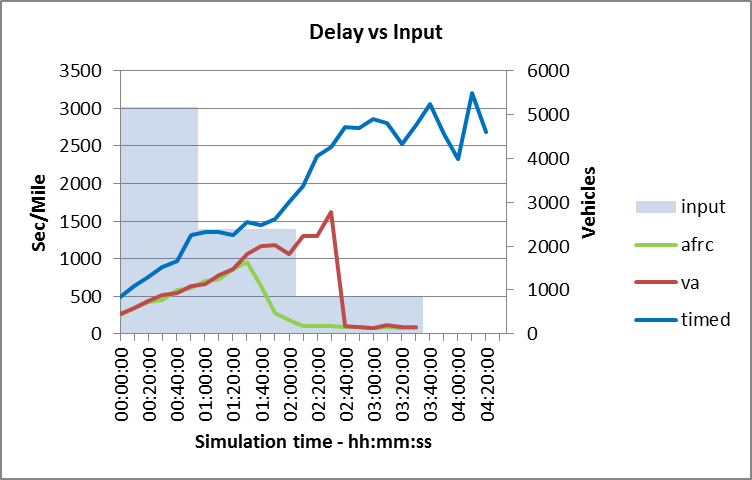


Figure .16: Experiment 4 Showing Traffic Input Volume

The number of “vehicles inside” the network is a useful indication of general network state but cannot be used in isolation to imply network efficiency. The comparison graph shown in Figure 6.13 may suggest that the AFRC controlled network is the least efficient as there are fewer vehicles being managed. However, if the same data is shown against vehicles waiting to enter, as shown in Figure 6.17 and Figure 6.18, the situation is clarified. The number of vehicles inside the network at any instant is equal to the difference between the number of vehicles offered, the number waiting to enter and the number that have completed a journey and have therefore left the network. As a mathematical proof:

Equ( ‑)

After rearranging Equ( 6‑3) to focus on the completed vehicles:

Equ( ‑)

It can be surmised that, given the same initial traffic volume, a smaller number of vehicles remaining inside the network when taken in conjunction with a smaller number still waiting to enter must be the result of more vehicles having left the network provided that the network is conservative. In terms of Queueing Theory, if the entire network is considered to be a two stage queue system similar to that depicted in , a smaller remaining input queue is only possible when the output stage is serviced at a higher rate if the internal buffer space remains constant throughout. In the case of this experiment, the input queue length is less and the internal buffer is also less therefore more vehicle must have passed through and left the system.

Taken in combination with the general equation for efficiency,

Equ( ‑)

it is suggested that the experiment version with the highest output level can be considered to be the most efficient. Substituting offered for input, completed for output

Equ( ‑)

Although this gives a general indication, the main limitation of Equ( 6‑6) is the delay in the queue waiting to enter and the duration of time that vehicles are inside the network. The completed term is a time delayed term relative to the offered term. Moreover, the waiting and inside terms have variable delay periods which are independent of each other. The step delays can be visualised by plotting the “Delay Time”, “Vehicles Inside” and “Waiting to Enter” profiles on the same axis for each system as shown in and .

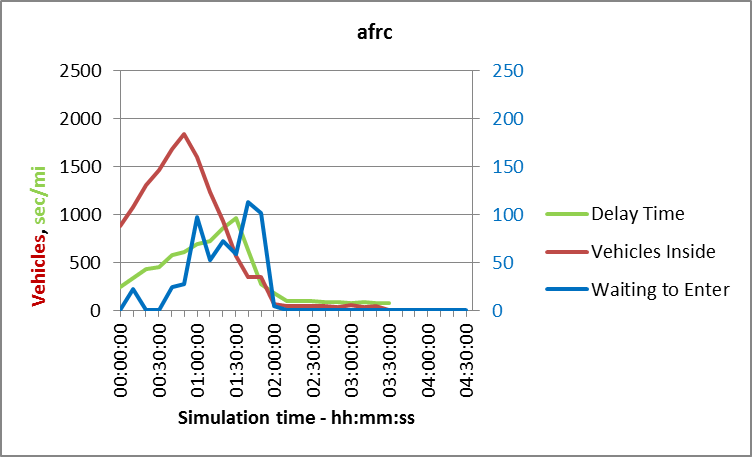


Figure .17: Experiment 4 AFRC Network Efficiency

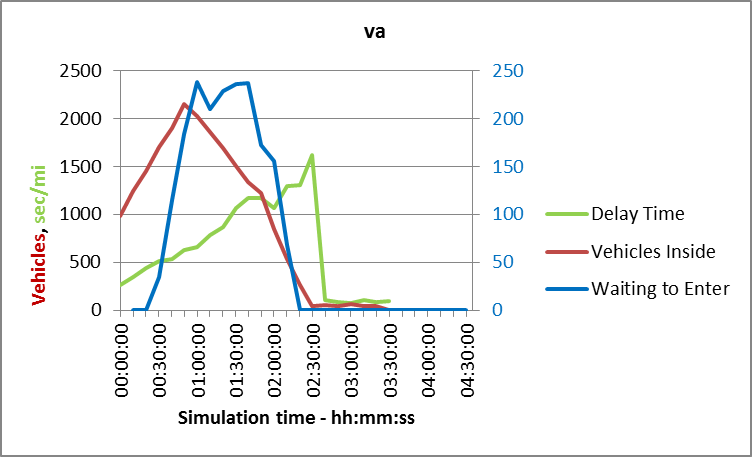


Figure .18: Experiment 4 VA Network Efficiency

It is interesting that the peaks for Delay and Waiting are reversed in the graphs, this is due to the superior vehicle removal rate for AFRC.

Conclusions that can be visualised from these graphs are that each of the components has a lower peak and a shorter duration. This is exactly as predicted in the introduction, smaller queues and shorter durations which will reduce pollution as well as travel time.

### Experiment 5 Results

Minimum green = 8 sec, Maximum green = 70 sec, Absolute maximum green=90 sec

|  |  |  |
| --- | --- | --- |
| Figure .: Experiment 5 Delay across Network | Figure .: Experiment 5 Stop Time | Figure .: Experiment 5 Travel Time |
| Figure .: Experiment 5 Vehicles Inside Network | Figure .: Experiment 5 Vehicles Waiting to Enter | Figure .: Experiment 5 Travel Distance |

### Experiment 6 Results

Minimum green = 8 sec, Maximum green = 40 sec, Absolute maximum green=60 sec

|  |  |  |
| --- | --- | --- |
| Figure .: Experiment 6 Delay across Network | Figure .: Experiment 6 Stop Time | Figure .: Experiment 6 Travel Time |
| Figure .: Experiment 6 Vehicles Inside Network | Figure .: Experiment 6 Vehicles Waiting to Enter | Figure .: Experiment 6 Travel Distance |

### Experiment 7 Results

Minimum green = 8 sec, Maximum green = 85 sec, Absolute maximum green=95 sec

|  |  |  |
| --- | --- | --- |
| Figure .: Experiment 7 Delay across Network | Figure .: Experiment 7 Stop Time | Figure 6.33: Experiment 7 Travel Time |
| Figure .: Experiment 7 Vehicles Inside Network | Figure .: Experiment 7 Vehicles Waiting to Enter | Figure .: Experiment 7 Travel Distance |

### Experiment 8 Results

Minimum green = 7 sec, Maximum green = 50 sec, Absolute maximum green=70 sec

|  |  |  |
| --- | --- | --- |
| Figure .: Experiment 8 Delay across Network | Figure .: Experiment 8 Stop Time | Figure .: Experiment 8 Travel Time |
| Figure .: Experiment 8 Vehicles Inside Network | Figure .: Experiment 8 Vehicles Waiting to Enter | Figure .: Experiment 8 Travel Distance |

### Experiment 9 Results

Average results from 10 simulations. Minimum green = 8 sec, Maximum green = 50 sec, Absolute maximum green=70 sec

|  |  |  |
| --- | --- | --- |
| Figure .: Experiment 9 Delay across Network | Figure .: Experiment 9 Stop Time | Figure .: Experiment 9 Travel Time |
| Figure .: Experiment 9 Vehicles Inside Network | Figure .: Experiment 9 Vehicles Waiting to Enter | Figure .: Experiment 9 Travel Distance |

### Experiment 5 – 9 Summary

The conclusions from the experiments are as anticipated; varying the max\_green and abs\_max\_green values had the greatest impact on the Timed results. Shortening the abs\_max\_green time by a modest amount could make the system more dynamic and tended to clear the traffic.

The worst condition for the Timed system was in experiment 6. The reason for this was that the signal timing was reduced to a level where there were greater total losses caused by frequent signal transitions to directions that had no demand.

In all cases AFRC resolved the network faster and with lower delay. The AFRC action results in rerouting traffic which may result in longer travel distances but in less overall time.

There are instances where the AFRC system shows a higher number of vehicles waiting to enter. The cause of this was identified as an ineffective location of the exit detector. This was discussed at the end of section .

## Experiment 10: Multidirectional Flow

Each of the previous experiments considered traffic flowing across multiple routes but with practical limitations. The main problem is that the flows are unidirectional, e.g. traffic travels from an Origin to a Destination but never returns e.g. a vehicle leaves “home” to go to “work” but does not return from “work” to “home” at any time. Besides being unrealistic, of greater significance is that each route is being tested in a single direction and only under a converging load profile. The significance of this was discussed in section where traffic density increases as traffic from many sources converges towards a single destination such as a major employer or the traffic density reduces as the traffic diverges towards multiple residential areas.

The regions were defined in section and are summarised below. The objective of this experiment is to test the ability of the AFRC system to manage converging and diverging traffic. The definitions are only very loosely based on a hypothetical layout to provide a level of context. The Origin / Destination in each case is shown as a single point, it is envisaged to be an area with multiple different destinations as shown in section ; the grid provides only the major access routes.

South-East. Commercial / industrial centre. The main destination for traffic during the morning rush-hour period and the main source of traffic during the evening rush-hour. Intended to represent a large volume of people travelling to and from a place of work. Some vehicles will travel to this area and then return to residential areas during the morning rush-hour, making the inverse journey in the evening. In terms of the testing, these are vehicles that are intentionally included to oppose the main tidal flow to force signal changes against the main route.

North-West. Residential 1. High density residential area where a high proportion of the workforce live. A main source of traffic in the morning and a main destination in the evening. In terms of testing this is part of a divergent network.

North-East. Residential 2. Lower density residential area, most vehicles from this area will not move from the commercial for the working day, mainly one way traffic flow in any period.

South-West. Residential 3. Small residential area. Minimal traffic in the main rush-hour periods, makes journeys to other areas during the off peak hours.

### Setup

The grid for this test looks very similar to the previous grid however it includes a change to the location of the exit detector near to the Origin-Destination centroid to allow detection of the queue entering the network.

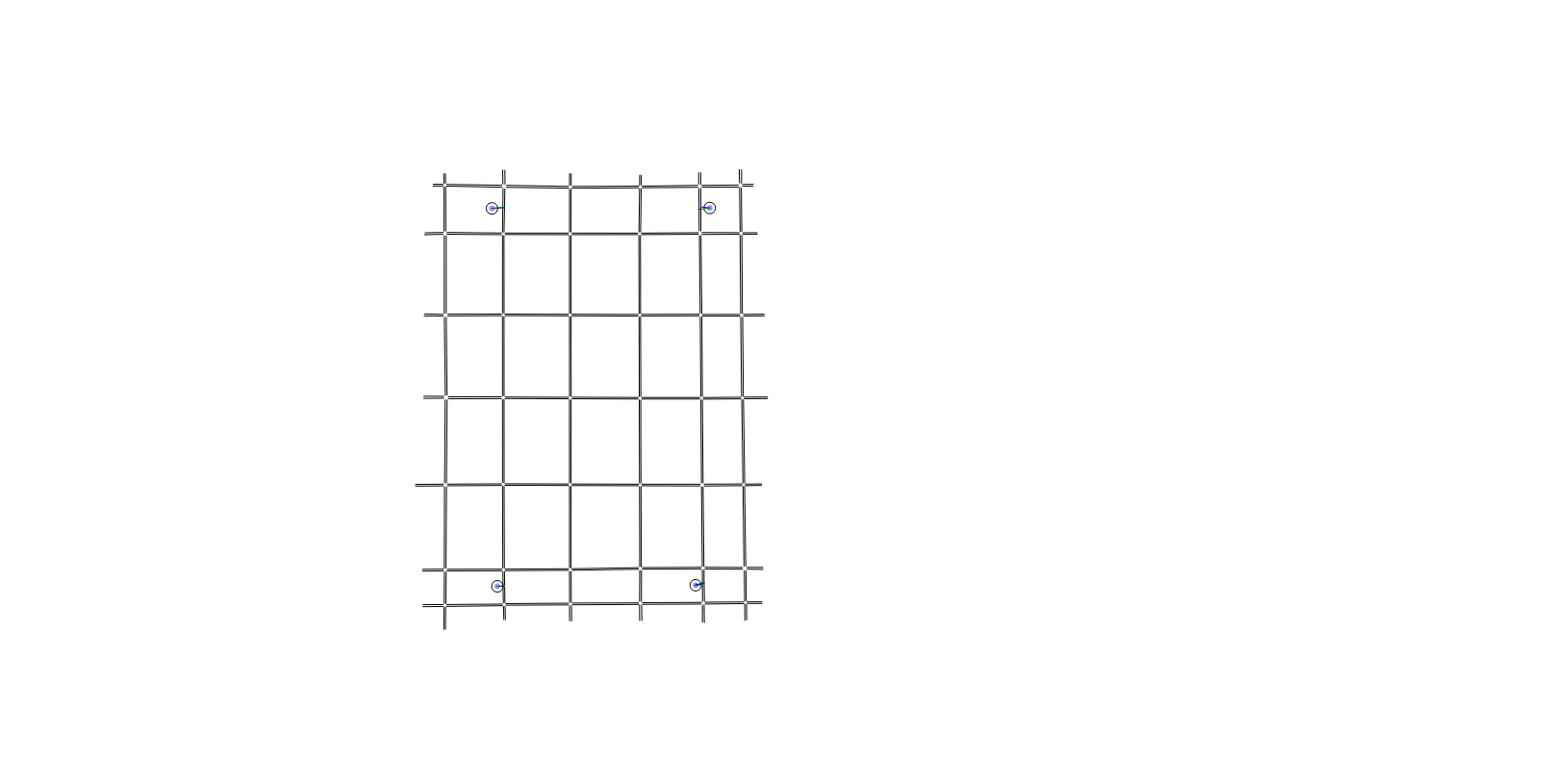


Figure .49: 7x6 Grid with Detector Update

The concept of the traffic profile in this experiment is that the major traffic flow will be towards a single destination area during a peak period at the beginning of the simulation. A period of low traffic flow follows then high volume divergent traffic, some minimal traffic and finally a quiet period.

Table .25: Experiment 10 Information Summary

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Experiment 10 | | | | | |
| Grid size | 7x6 | | Version | 6 | |
| Traffic Flow  (hh:mm) | 00:00-01:00 | 01:00-02:00 | 02:00-03:00 | 03:00-04:00 | 04:00-4:30 |
| Peak to commercial | Modest to-from commercial | Peak from commercial | Minimal residential | None |
| File | rectangle oversize 7 x 6 -realflow – v6.ang | | | | |
| Controller | tls-controller-rectangle-oversize-7x6-v7.py | | | | |

Table .26: Experiment 10 Traffic Flow - Peak to Commercial

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id:name | 775:  South-East | 783:  North-West | 1446:  South-West | 1449:  North-East | Total |
| 775: South-East | 0 | 100 | 5 | 75 | 180 |
| 783: North-West | 2000 | 0 | 0 | 0 | 2000 |
| 1446: South-West | 1000 | 0 | 0 | 0 | 1000 |
| 1449: North-East | 2000 | 0 | 0 | 0 | 2000 |
| Total | 5000 | 100 | 5 | 75 | 5180 |

Table .27: Experiment 10 Traffic Flow – Modest to-from Commercial

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id:name | 775:  South-East | 783:  North-West | 1446:  South-West | 1449:  North-East | Total |
| 775: South-East | 0 | 10 | 40 | 40 | 90 |
| 783: North-West | 10 | 0 | 5 | 5 | 20 |
| 1446: South-West | 40 | 5 | 0 | 20 | 65 |
| 1449: North-East | 40 | 5 | 20 | 0 | 65 |
| Total | 90 | 20 | 65 | 65 | 240 |

Table .28: Experiment 10 Traffic Flow - Peak from Commercial

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id:name | 775:  South-East | 783:  North-West | 1446:  South-West | 1449:  North-East | Total |
| 775:South-East | 0 | 2000 | 1000 | 2000 | 5000 |
| 783:North-West | 100 | 0 | 0 | 0 | 100 |
| 1446:South-West | 5 | 0 | 0 | 0 | 5 |
| 1449:North-East | 75 | 0 | 0 | 0 | 75 |
| Total | 180 | 2000 | 1000 | 2000 | 5180 |

Table .29: Experiment 10 Traffic Flow – Minimal Residential

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id:name | 775:  South-East | 783:  North-West | 1446:  South-West | 1449:  North-East | Total |
| 775:South-East | 0 | 0 | 0 | 0 | 0 |
| 783:North-West | 0 | 0 | 5 | 5 | 10 |
| 1446:South-West | 0 | 5 | 0 | 20 | 25 |
| 1449:North-East | 0 | 5 | 20 | 0 | 25 |
| Total | 0 | 10 | 25 | 25 | 60 |

Table .30: Experiment 10 Traffic Flow - None

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id:name | 775:  South-East | 783:  North-West | 1446:  South-West | 1449:  North-East | Total |
| 775:South-East | 0 | 0 | 0 | 0 | 0 |
| 783:North-West | 0 | 0 | 0 | 0 | 0 |
| 1446:South-West | 0 | 0 | 0 | 0 | 0 |
| 1449:North-East | 0 | 0 | 0 | 0 | 0 |
| Total | 0 | 0 | 0 | 0 | 0 |

### Results

The results from this experiment show the effects of converging and diverging traffic flow. The first result is Delay across the network which shows a very similar profile to the previous experiments with AFRC delay initially comparable to VA and clearly better than Timed. The peak delay for AFRC is lower than for VA and the delays reduce sooner. The second half of the graph shows a flat delay response for each system, with VA and AFRC displaying significantly lower delay than Timed.

The traffic profile is however different from previous experiments A comparison is given by and . Note that the traffic profile for experiment 10 is two peak flows and is given as absolute values; some vehicles are moving in directly opposite directions, vector addition would remove these from the results. The traffic flow at any centroid has apparent positive (e.g. inbound) and negative (e.g. outbound) flows; the centroid flow for the commercial area is shown in . In previous experiments the vehicles always moved in the same direction and the flow reduced in stages.

The similarity to the previous results is due to AFRC and VA resolving the commercial area inbound flow before the outbound flow begins; AFRC resolves the rush-hour peak around 40 minutes after the first peak completes, VA takes approximately 30 minutes longer. Timed does not resolve the inflow before the outflow begins and therefore has two contending flows to manage simultaneously.

It may be assumed that the delays for the Timed controller should indicate an increasing step change in the second half of , i.e. when the return rush-hour begins. The increasing number of vehicles that are waiting to enter, as shown in , account for the discrepancy. The conclusions section of experiment 3 (section ) identified that the input capacity for the Timed system is less than 1800 veh/hr; in section calculates a per-direction rate of <840 vehicles/hr given the abs\_max\_green value of 70 sec (cycle time ≈2 x green time with 50%split) or 1680 vehicles/hr maximum for two directions simultaneously. Traffic is trying to enter the network at a rate of 5180 vehicles/hr and is therefore forced to queue outside.

The waiting to enter graph, , also shows that traffic is queueing for both VA and AFRC which is due to a similar capacity restriction. Junctions with either VA or AFRC do not have a fixed approach capacity due mainly to the variable cycle time. For these systems, the maximum cycle time is the same as for Timed however this will only occur when there is no traffic flow to trigger a signal change (or all detectors have simultaneously failed); this is unlikely to occur and is considered to be a trivial solution when it does as there will be not traffic to manage. The minimum cycle time occurs when there is traffic flow in one direction or additionally, in the specific case of AFRC, traffic spillback to the exit of the previous junction. In this case, the absolute maximum green time (incorporating the single yellow) will apply in one direction, followed by red. The red has a duration of the combined minimum green for the conflicting flow and a minimum 2 second red/yellow (amber). Using the abs\_max\_green time of 70 seconds, the min\_green of 8 seconds and assuming no all-red time, the cycle time for a simple system, as depicted by would be 80 seconds. The split time, according to Equ( 3‑23), is given by

Equ( ‑)

The split% indicates the percentage of time that a green signal is shown and can be used to calculate the actual capacity of a junction approach assuming that there is a sink capacity which is equal to or greater than the approach rate. In this case the capacity per direction will be

Equ( ‑)

or approximately 3150 vehicless/hr for traffic that is permitted to enter the network in two directions simultaneously from a centroid.

For VA and AFRC, the cycle time can increase if there are additional stages however, AFRC can skip stages that are not required due to empty approach queues or where the exit is blocked. AFRC is able to manage queueing traffic better than VA due to the additional information that is provided by the addition of the junction exit detector. This is advantage is shown by the lower number of vehicles in the second return rush-hour stage of the “waiting to enter” graph.

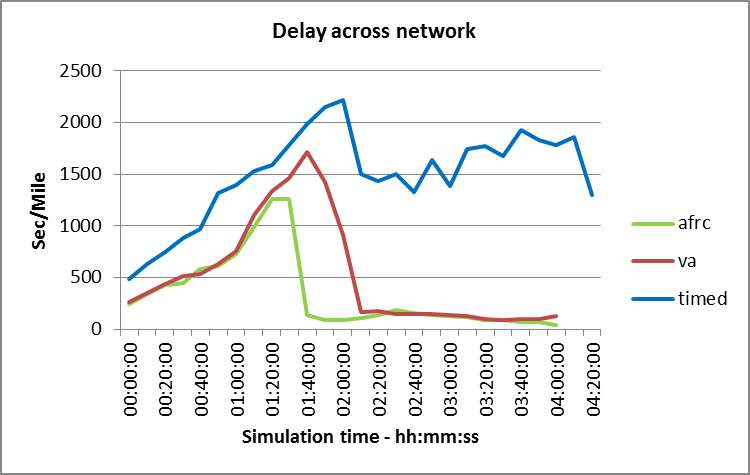


Figure .50: Experiment 10 Time Delay across Network

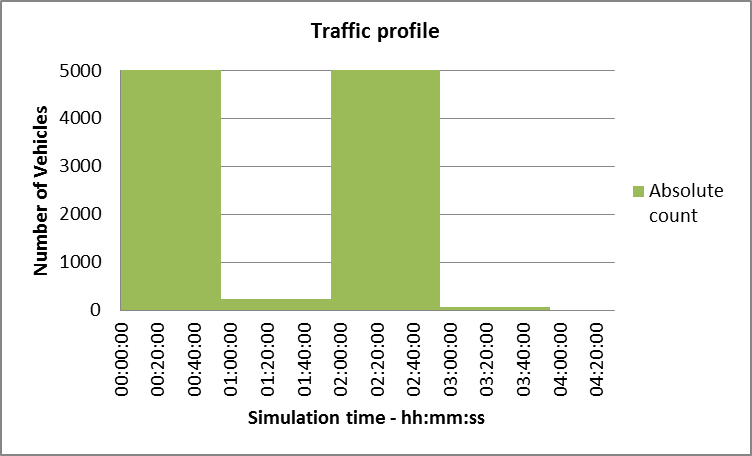


Figure .51: Experiment 10 Absolute Traffic Profile

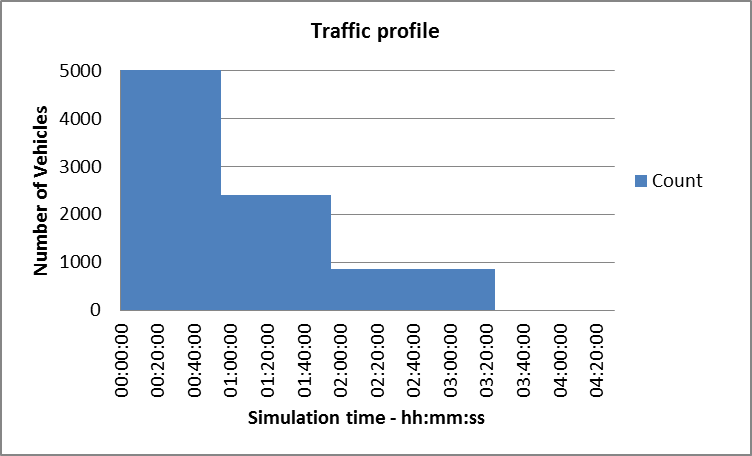


Figure .52: Experiment 3 - 9 Traffic Profile

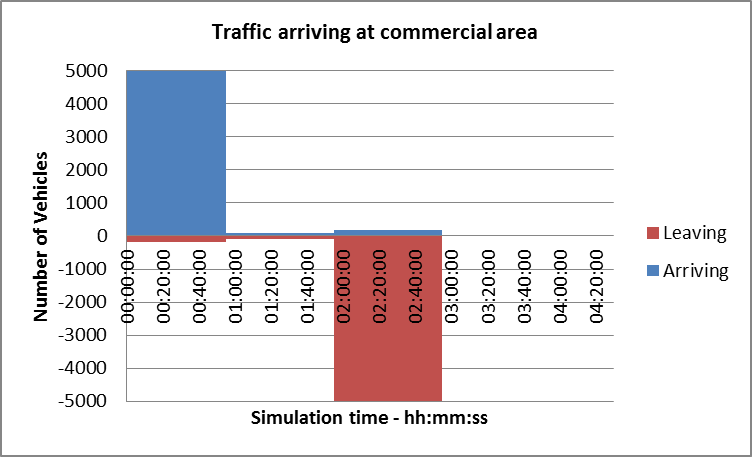


Figure .53: Experiment 10 Commercial Area Traffic Flow

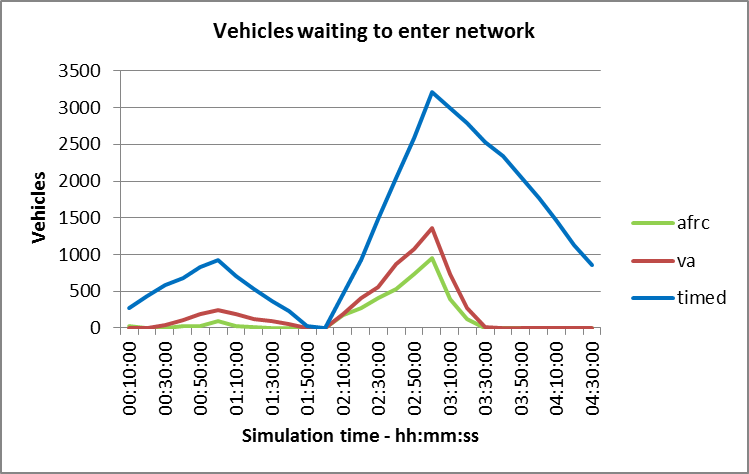


Figure .54: Experiment 10 Vehicles Waiting to Enter Network

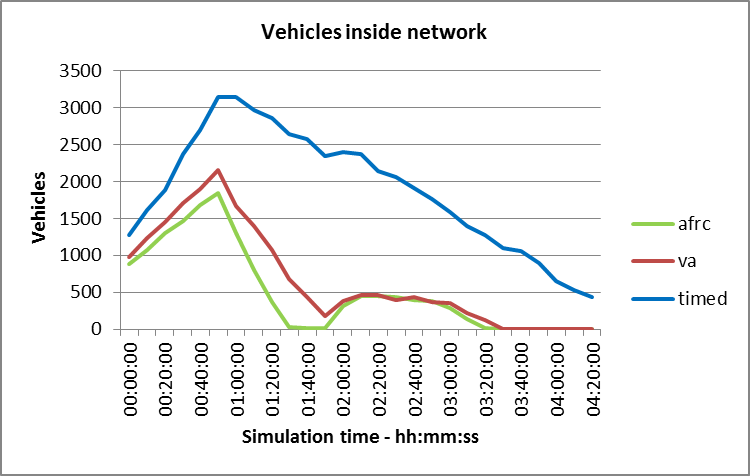


Figure .55: Experiment 10 Vehicles Inside Network

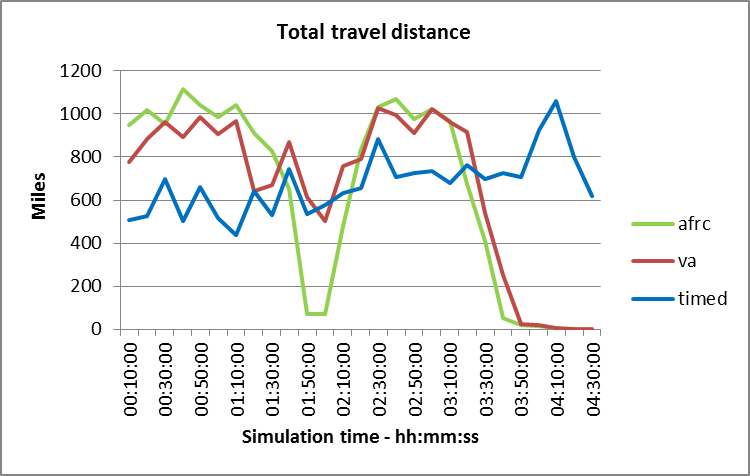


Figure .56: Experiment 10 Total Travel Distance

### Summary

These experiments were designed to investigate whether or not the AFRC system is able to reduce congestion in terms of numbers of vehicles involved and the time duration of queueing.

The delay across the network graph does not indicate that AFRC has a significant improvement over the VA system when the traffic flow is diverging. However, because less vehicles are queued according to the “waiting to enter” graph, , more vehicles must be entering the network with AFRC. Also, the “vehicles inside” graph, , shows similar numbers of vehicles with AFRC and VA but this means that the direct consequence of more vehicles entering is that more have completed their journeys and left the network. This is conclusion is supported by the number of vehicles inside falling to zero sooner for AFRC than for VA.

Waiting to enter shows that the entry point from the “commercial area” is a focal point where the traffic density is highest. This was not being managed by the proposed control system. In a practical road network, this would represent the final stage of the private land e.g. the car park exit. It is possible to include exit detection but this would be a commercial consideration with legal agreement outside of the scope of this work. Traffic is being managed effectively after joining the main roadways, as confirmed by the time delay and vehicles inside graphs ( and ).

# Conclusions and Future works

## Conclusions

The aim of this work was to identify a means to reduce the duration and negative effects of urban congestion through the development of a novel traffic control method. The work has successfully identified and developed an algorithm termed Available Forward Road Capacity (AFRC) that can achieve this aim with the use of an additional stage of vehicle detection that is not present in existing controllers. The AFRC system directly measures the vehicle capacity of a downstream link to determine the viability of forwarding traffic; existing systems estimate this capacity and adjust the junction timing according to the estimate. Existing systems may underestimate the downstream capacity causing inefficient operation due to early signal transition or overestimate the capacity causing spillback blocking and preventing practical control action. In either case, the system operates below maximum capacity which needlessly extends queue and congestion duration. The accuracy of the estimate can be affected by vehicle type, size, weight and by driver response. AFRC measures the actual capacity and takes immediate appropriate action. The detection is unaffected by the attributes of either the driver or vehicle and is therefore accurate at all times.

A range of existing detection methods can be used as a source for the AFRC control algorithm. The algorithm can be implemented as a part of, and in conjunction with, existing signal controllers. The use of existing detection systems means that they are practical, available, affordable, have known output characteristics and did not require development as a part of this project. The AFRC control algorithm affects traffic flow only when the controlled junction is experiencing saturation levels of traffic, where existing traffic controllers are no longer effective.

Simulations have been conducted using Aimsun, a highly reputable commercial traffic simulation tool that is utilised by transportation planners in many cities across The World. The novel control algorithm was developed and coded using Python and applied using the Aimsun API. The coding included a Vehicle Actuated (VA) system implementation similar to MOVA, a very common signal control system, and a purely time activated control system.

Road layouts and traffic profiles were defined in order to test the response of AFRC algorithm in comparison to the VA and Timed systems. The benefit of detecting vehicles at junction exits using AFRC was proven in all cases to reduce the numbers of queued vehicles, the length of time spent in a queue and the overall duration of the congestion period.

Reducing the number of vehicles queued, the queueing period and overall duration of the congestion will reduce the total amount of energy required by the system which will reduce fuel waste, pollution and environmental damage. Reducing the overall energy requirement will reduce pollution regardless of the type of fuel used, savings can be attributed to hydro-carbon fuel burn, electrical power generation and increased battery / fuel cell lifetime due to reduced charging cycles. Improving battery lifetime reduces environmental damage from chemical and material use during production and from end of life disposal.

Alternative methods of congestion reduction including Connected and Autonomous Vehicles (CAV) have been proposed elsewhere but will require many years before full or even significant adoption. They may also require significant enabling infrastructure changes that will be slow to implement in large but busy metropolitan areas and potentially unaffordable in the short-term for smaller Local Authority budgets. Until these changes are implemented and proven to be effective, changes can be made to reduce needless congestion and environmental damage caused by the inefficient use of traffic control signals.

It is speculated that CAV will ultimately replace the need for traffic control signals, however there will still be a need for independent arbitration of contented junction space. The conceptual ideas proposed by this work will still form the basis for the efficient utilisation of the contended space regardless of the control implementation

## Limitations of AFRC

The proposed AFRC system appears to have many benefits over existing traffic control systems however it should be noted that there are also limitations.

First and foremost, AFRC is a software extension to an existing traffic control system and not a replacement. The reason for this being that AFRC is specifically intended to adjust signal sequencing and timing only under conditions when the junction is over saturated, the existing signal algorithm being unable to maintain active control and having effectively reverted to a timed or preconfigured sequence.

AFRC requires the addition of a detector in the exit of each controlled junction. It is not necessary to implement this at every junction simultaneously and the detector does not need to be the same type at every junction but the overall control system does rely on this for for functionality. In mitigation, the detector cost should be minimal and many SCOOT junctions already have a detector that could be utilised although it is currently attached directly to the downstream junction.

Detectors are necessary for the entire traffic control system to function regardless of whether AFRC is implemented however the AFRC would be compromised if there was a failure of the exit detector. In mitigation, if the exit detector fails, the junction would simply revert to a standard junction without AFRC enhancement.

The AFRC algorithm relies on having the ability to influence the next downstream control point, this is only possible where the control point is an active system i.e. a traffic light and not a “Give Way” restriction. Traffic engineering would be required to resolve spillback problems caused by road layout before AFRC could improve traffic flow. The required traffic engineering could necessitate the introduction of new traffic light signals which would incur additional cost and reduce traffic flow at any time other than when the junction is overloaded; an analysis would need to be undertaken to assess the overall benefit.

## Future Works

Throughout the duration of this research, multiple alternative approaches aimed at resolving the problems of traffic and congestion management have been either proposed or further developed. This final section discusses some of these proposals, explaining why they were not included in the main body of this work and identifying opportunities for continuing research in fields related to future transportation systems.

### Autonomous Vehicles

There is speculation that Autonomous or, more precisely, Driverless vehicles will resolve all of the problems of congestion and prevent accidents. This is a recurring subject that has failed to meet expectations each time it arises. A work by Kröger [122] states that the probability of autonomous vehicles has been 20 years into the future for very nearly 100 years. Kröger lists many of the early attempts at self-driving vehicles with examples from as early as 1921. Amongst other details, the work includes a discussion regarding an educational film production from GM in 1935 entitled “The Safest Place”. In the discussion, Kröger identifies a failing of the work insomuch as it fails to recognise the vehicle itself as being a risk; that an accident can still occur without driver error as a cause. This appears to be a common problem in many of the research papers over the years.

Kröger [122] continues to discuss the conceptual ideas of wire following, line following from the 1950s, vision guidance systems of the 1970s and the later self-contained systems of the 1990s. The point is made that there is a transition from automated roads to automated vehicles. The work also discusses benefits that have been gained from the developments including cruise control and self-guided extra-terrestrial exploration vehicles such as “Moon Rover” and the comparison between the capability of science fact and the imagination of science fiction.

Autonomous vehicles will not resolve the congestion problem. In the simplest terms while there are more vehicles than junction capacity there will be queueing. As stated in the justification of this research from para : congestion typically occurs when traffic queues are not resolved within a reasonable period. In the language of queueing theory, queues occur when the number of customer arrivals exceeds the service rate of a control point. Using this definition, in order to resolve urban congestion there are two alternatives: increase junction capacity beyond the level of customer arrivals or reduce the number of customer arrivals below the junction capacity. Driverless vehicles achieve neither of these alternatives and so cannot mathematically resolve congestion.

The terms Autonomous and Driverless are frequently used interchangeably however it should be clear that practically all road vehicles in use today are autonomous; the driver / vehicle unit is only aware of its own location, environment and intention. Expectation of interaction with other vehicles is determined through experience, moral obligation and legal guidance but the ultimate interaction decision are based on assumption of the vehicle response to a control input and a prediction of the actions of all other autonomous vehicles in the local area. Accidents between vehicles happen when the assumption or prediction is incorrect or an undesirable control input is applied. Autonomous vehicles cannot fully address these situations.

Driverless vehicles may be autonomous, only aware of their own environment and intentions, or connected and aware of the environment and other vehicles through the use of vehicle to infrastructure (v2i) and vehicle to vehicle (v2v) communication.

In order to achieve the safety claims or expectations of autonomous vehicles slower movement and greater decision time are required. This will increase journey times meaning that vehicles will be on the roads for longer time periods, vehicle density will increase for the same number of vehicles to complete less journeys in a given time and as a result congestion will increase in every definition of the term.

Justification of congestion claims frequently made by autonomous vehicles requires a social change towards “Mobility as a Service” (MaaS) where the concepts of pay-per-use, shared ownership, car-pooling and shared occupancy are considered. These are options that are currently available and in use, unless there is a massive change in social attitude, there is no reason to expect the current levels to change. A method of promoting the change of use is through road pricing and cost of ownership.

### Connected Vehicle Priority

Connected vehicles are expected to communicate with each other and with infrastructure. There are technical considerations which are outside of the scope of this work but there are some simple logical concepts.

Based on the assumption that there will be queueing even with the advent of autonomous and connected vehicles, it is further assumed that the queueing will still be tidal and asymmetrical at each junction. In order to minimise the overall congestion, vehicles will need to negotiate flow ratios per approach based on arrivals, queue lengths, times and days. To ensure fair allocation vehicles would not be able to decide on their own priority in comparison to other road users and so some form of independent arbitration system will be required. Without this, there is a potential for two or more vehicles to prioritise themselves independently and either crash or reach an impasse resulting in gridlock. Current traffic light controllers are independent arbiters capable of determining and applying load allocation for specific junctions with unique physical and temporal characteristics and are likely to remain in service if connected vehicles become the norm. The current system uses an observable indication, it may be possible to modify this to be communicated by wireless methods but this would rely on every junction user (car, bicycle, pedestrian, etc) being connected and with zero equipment failures. It is more likely that the connected vehicle system would either read the optical signal or will operate in parallel. This makes it highly probable that the current traffic signal and controller will remain in service for the future. The conflicting directional requirements, queueing and congestion will still benefit from the AFRC algorithm.

Other points are that an individual vehicle that prioritises itself over others would be the equivalent of running a red light in todays road networks. This would be impossible to police without independent arbitration.

Solutions for connected vehicles include adjusting approach speeds with the aim of never having a vehicle actually stop. As shown in section , vehicles in motion occupy more space than vehicles at rest. Having the same number of vehicles with larger inter-vehicle gaps means that the queue will occupy more physical space, extending the congestion area and interacting across more junctions. The queue will appear less dense but the psychological effect would be the same.

Modern combustion vehicles are equipped with “auto-start” systems that temporarily turn off the engine while the vehicle is stationary in order to reduce fuel burn, increase fuel efficiency and reduce environmental pollution. Slow moving queues of connected vehicles would use not be able to utilise this system and would as a result cause more environmental damage. Slow moving vehicles are typically in low inefficient gears with increased fuel use, this is the case even for many electric and hybrid vehicles. There are statements that there will be increased fuel use from a standing start however there must be a crossover point from very slow moving in low gears over a long period and approaching in a higher gear and stationary for a shorter period with no drive energy.

### Reduced Headway and Reaction Speed

The concept of headway and breaking distance was discussed in section . The section shows the minimum desired safe distances between vehicles and identifies that the currently accepted inter-vehicle gap of 2 seconds does not provide enough distance to stop even applying a 1G deceleration rate. The force of 1G is extremely high, Gates et al. measured an extreme maximum of 0.78G in their study [89]. The 1G stopping times table from is reproduced as .

It is a common statement that the main advantage that autonomous systems have is the speed of reaction in comparison to human capability. This may be highly misleading. The figure quoted for human reaction time includes a proportion of decision time. The decision time cannot be separated from human reaction times but is easily determined in computing systems and is commonly ignored. Evidence to support this view can be found in the NTSB interim report for the fatal incident between an Uber test vehicle and a pedestrian / cyclist at 21:58 on Sunday, March 18 2018 [123]. The report states that the deceased pedestrian was observed 6 seconds before impact at 43mph, classified as unknown object, then vehicle and finally bicycle. The system determined that action to “mitigate a collision” 1.3 seconds before impact. The vehicle did not make an automated response and so it is not possible to obtain a time for this. What is clear from the report is that the decision process took 4.7 seconds and the vehicle recognised that at 43mph a collision could only be mitigated, not avoided, in 1.3 seconds. This agrees with the braking time data identified by .

Table .1: Stopping Times Based on 1G Deceleration

|  |  |  |  |
| --- | --- | --- | --- |
| Speed  (mph) | Calculated times (sec) | | |
| reaction | braking | total |
| 20 | 1.47 | 0.91 | 2.38 |
| 30 | 1.47 | 1.37 | 2.84 |
| 40 | 1.47 | 1.82 | 3.29 |
| 50 | 1.47 | 2.28 | 3.75 |
| 60 | 1.47 | 2.73 | 4.20 |
| 70 | 1.47 | 3.19 | 4.66 |

In addition to the decision time, human – machine interfaces (HMI) typically require a degree of multiplexing. This translates to the time required for physical movement e.g. moving from accelerator pedal to brake pedal, pressing the pedal and having this transferred by compressible fluid into actual braking action. Automated systems will typically have an actuator dedicated to each function and therefore no physical movement is required.

In summary, it is not realistic to expect that autonomous systems can operate with reduced headway and as a result will not use less road space or reduce congestion.

### Connected Vehicle Platooning

Similar to reducing headway, the concept is that a number of vehicles, typically HGV, will form a type of road train with distances maintained by vehicle systems rather than physical link. This is more suited to inter-urban routes and is therefore out of scope for this work however there are some comments that can be related. First is the inter-vehicle gap, this will need to be more than for gap for a car which, according to , is 176 feet (54m) at 60 mph. Keeping this gap free from other vehicles would be difficult to maintain and so following vehicles would be forced to move back, disconnecting from the platoon.

As implied by Kröger, accidents can be caused by mechanical faults [122] and these will be unavoidable with short gaps. The mitigation would be that with very small gaps, the closing speed would be minimal and the accident, although large, is unlikely to be directly fatal.

The need for platoons is debatable, it implies that more goods than can be carried in a single vehicle need to be transported from a single source to a single destination at the same time otherwise the platoon is not viable. If the platoon is purely a motorway event then it will not be possible for the driver of any vehicle to exit the vehicle with the platoon in motion. If the driver remains in the vehicle there is no cost advantage for the haulier; the risk of accident has increased with no mitigating cost reduction, in fact the cost of purchasing and fitting platooning controls could be excessive.

Finally there is the question of forming practical platoons. Either the “front” vehicle would have to extend delivery time by travelling below its maximum speed to allow other vehicles to “catch up” or following vehicles would have to exceed legal (and governed) speed limits. This would need to continue until the desired platoon length is reached.

### Road Pricing, Cost of Ownership

Increased cost of ownership is an option as a means to reduce the numbers of vehicles on the roads and hence congestion. The main problems with this concept is that personal transport will become a benefit only affordable to the more affluent society and no longer available to the lower paid working classes. As a result, the working classes would not be able to afford to live away from employment areas or regions with very good public transportation. This would be a retrograde step in socio-economic mobility and have significant impact on local and national economy.

If increasing the cost of ownership does not result in increasing wages to compensate, then it is likely to reduce car ownership. Reducing car ownership will reduce demand, reduced demand will reduce profits and either reduces purchase price leading to increased purchasing or requires a reduction in production. The economic benefit of car production in many countries is exemplified by the countries that implemented “car scrappage schemes” as attempts to stimulate national market growth during the global economic downturn in the late 2000s. The schemes offered manufacturer and government funded discounts on the purchase of new vehicles when vehicles above a certain age were taken out of use in the same transaction. Research suggests that the schemes implemented in many European Countries in 2009 added as much as 0.26% to European GDP, and was one of the main reasons for the Eurozone recession recovery in that year [124]. Schemes had been offered previously with the primary goal to remove older, less efficient vehicles with higher emission levels [125]

Countries that are considered to be cycle and pedestrian friendly include Denmark and The Netherlands. Amsterdam and Copenhagen are frequently discussed as being examples of how transportation should be managed and that they require ever greater spaces for bicycle parking. Another significant similarity is that the two countries share is a lack of significant automotive industry. Reduction in the numbers of cars sold globally would have little, if any, negative impact on either countries economy. This is evidenced by the European Car Scrappage schemes during the European economic downturn in 2009. There were 13 EU member state schemes in that year, 10 of which were introduced primarily in response to the economic crisis. There was no scheme implemented in Denmark and the scheme in The Netherlands was primarily focused on environmental benefits [126]. Cars in either country represent imports and therefore have little bearing on employment. This removes the pressure on the government to maintain car use specifically to sustain production and hence protect jobs and the economy.

### Shared Ownership, Pay-Per-Use

In order to reduce congestion, the number of vehicles can be reduced. One option for this is to avoid private ownership and instead use a shared ownership or pay per use model

The “pay per use” concept already exists in the form of a taxi or private hire vehicle, which also resolves the mobility problem for those unable to drive. Driverless pay-per-use is suggested as a cheaper alternative to the Taxicab as there is no driver wage to pay. However, the calculation of cost needs to include the fact that the driverless vehicle will cost significantly more than the equivalent regular car. The cost implication is simple, the driverless version will be based on the regular version, with the possible exception of the minor points such as the steering wheel and instrument cluster, nothing will be removed and so there will be no cost saving. In addition to the regular model, a navigation system, suite of sensors (lidar, radar, vision, ultrasonic), high performance computing system and a v2x (vehicle to anything) communication system will need to be added. Regardless of development and mass production cost reductions, these components are still in addition to the basic cost and so the car will never be cheaper than the regular version. The next factor to consider is that the control system will have a finite life which is much shorter than the mechanical life of an existing taxi; driverless versions will need more frequent replacement or major upgrade. This assumes that there is no manufacturer added cost upgrade during the useful lifetime of the vehicle. Finally, if the vehicle is replaced due to obsolescence then it has no resale value and is a total loss.

The next often quoted supposition is that the vehicle will be able to carry passengers 24 hours per day and 7 days per week therefore will have a faster return on investment. The problem with this concept is that there is no demand for 24x7 transportation. Many private vehicles travel from a residence to a workplace where they sit in a carpark for the day then travel home and sit overnight at the residence. There is no need for these vehicles to be used at any other time. However, there is a need for hundreds or thousands of vehicles to make a similar journey at a similar time each day. In terms of pay-per –use much like taxicabs today, there would not be enough conveniently placed during the rush hour demand and far too many during the quiet period of the day.

The most obvious discrepancy between concept and reality is that the same number of journeys and journey miles are required to move the same number of people between destinations. If the number of vehicles is reduced, many will need to make multiple journeys, carrying passengers in one direction but not in the other. This would reduce congestion due to the reduced number of vehicles however the number of vehicles would create solid rather than tidal flow, with more intense interaction at each junction; queueing and congestion would actually increase while the probability of completing a journey on demand would reduce.

A further concept for driverless pay per use is that it would avoid the need for car parking however the vehicles still need to be somewhere when not in use. They could continue to circulate using fuel and occupying road space or they can park. Taxis in Derby are yellow and easily recognisable in Figure 7.1 and Figure 7.2. Regardless of definition, the vehicles have to occupy car parking at some stage.

Car-pooling or shared ownership is only viable when there is absolute trust between the co-owners and certainty that the vehicle is not required by more than one of the owners at any one time. Discord could ensue if the vehicle was consistently unavailable due to it being in use or requiring servicing or repair. To avoid this situation there should always be more availability than need which means that the vehicle would still be unused for large parts of the day.



Figure .1: Google Image of Derby Railway Station



Figure .2: Google Image of “The Spot” Derby

Car-sharing is a concept that should reduce journeys but people do value their own space and privacy. Car-sharing requires multi-party commitment and can be physically, temporally and emotionally restrictive. These may be reasons why car-sharing is rarely popular even when incentivised by mass employers.

### Big Data and Traffic Predictions

The objective of Big Data analytics systems is to determine trends and influences from datasets that are so large that relationships are obscured by the mass of data. It is considered possible that traffic flow could be predicted if enough data was collected that includes individuals, travel and routes. The Big Data concept also thrives on seemingly random data being used to identify a particular event for an unrelated system. There have been several studies that have been conducted in an attempt to prove this possibility. Xu et al. proposed a system that would collect massive amounts of historic data collected from existing roadside instrumentation which is then processed in order to make a traffic prediction for the future [127]. The work concludes that they can produce a prediction of traffic conditions which is 90% accurate over the single 3.4 mile section of freeway. The accuracy is obtained by measurement of the actual traffic conditions at the predicted time. This prediction works but it is not comparable to the prediction necessary for accurate control of traffic flow in an urban environment. For the prediction method, the vehicles are already in the system and cannot be removed; with the urban controller, vehicles may enter the system at some undefined time and cannot be accounted for at the precise point of leaving “home”. As discussed in section 2.3.5, in a controlled road network, queueing begins when a single vehicle (or fraction of a vehicle) exceeds the capacity of a junction. The stated 90% accuracy may be acceptable for a bulk (macro) flow for road planning but is unlikely to be acceptable for signal timing.

The problem of initial conditions i.e. the indeterminate journey start time is similar to that of Chaos theory discussed in section 2.4.1. Prediction of a future event requires precise knowledge of all possible variables at some initial time t0. It is not possible to predict the precise time that a journey will begin and so t0 cannot be identified in advance. This becomes highly significant when considering that the journey for every individual will have a unique t0 which is not synchronised to any other journey or the signal controller. The starting time of any journey is infinitely variable even though this is likely to be within between the bounds of a few minutes, this will affect the road position in the converging system discussed in section 4.4 with the subsequent effect of changing the arrival at the control signals in terms of position in queue and current signal cycle. This fits precisely with the quote taken directly from the Lorenz paper [47] abstract: “*For those systems with bounded solutions, it is found that nonperiodic solutions are ordinarily unstable with respect to small modifications, so that slightly differing initial states can evolve into considerably different states.*”

A comparison can be drawn to the idea of using Big Data analytics to determine where and when housing should be built to accommodate a local population. Big Data may be able to achieve this as a “bulk response” however it will not be able to predict precisely which brick will be placed in which order and the exact time it will be laid; it will only predict the buildings will be needed. For urban traffic, Big Data could determine future road capacity but not the precise time or order that individual vehicles use it.

## Potential Research Fields

There are many alternative proposals to resolve congestion however history, elementary physics and simple economics can be used show that congestion is a complex problem and will not be resolved while there remains a need to transport goods or people. The following three areas are being considered for future research.

### Connected and Autonomous Vehicles

Optimistic views propose that driverless vehicles will be available in the early to mid-2020s although there is competition to be the business making a claim to be the first. If the middle of the next decade is representative and actually delivered on time, there is still a 10 year intervening period where congestion effects need to be reduced. The technical capability and availability of driverless vehicles do not mean commercial viability, legal compliance, social acceptance or compatibility with existing vehicles, existing controls or vehicles from alternative manufacturers. It also does not mean that there will be, or even can be, an immediate switch over to the technology, there are unfounded predictions that any change to autonomous vehicles could take up to 40 years and involve several technology and cultural refresh periods. The proposal of driverless vehicles are often linked to other ideas to justify how congestion reduction will be achieved, this concatenation of ideas clearly confirming that it is not universally accepted that driverless vehicles will have a positive impact. Many of the supporting schemes, such as car-sharing, could be implemented in isolation but they require a cultural shift as an enabler and are reliant on a paradigm shift such as driverless vehicles as a catalyst for change.

### Traffic Signals Communication

The CAV concept may necessitate replacement of conventional traffic control signals, however there will still be a need for independent arbitration of contented junction space. Future controls may communicate more information and/or use different methods for communication. There is scope for research into the carrier system, messaging protocol and message security/confirmation.

### Prediction Methods

Predictive controls based on the ideas of Big Data rely on precise timing, knowledge and control, these are not realistic when the primary subject is a free thinking unique and independent processing system such as a human. Transportation is however a complex subject that is not limited to individuals within the confines of urban environments. There is extensive scope for Big Data analytics in the field of predicting requirements for mass goods transportation.

### Economics and the need for change

Socio-economics and venture capitalists will determine acceptable levels of congestion in terms of acceptable return on investment, whether the investment is money, time or perceived social / moral standing. While this is clearly not a technical field, the need for change is a more powerful force than technical capability alone.

# References

|  |  |
| --- | --- |
| [1] | M. H. Lighthill and G. B. Whitham, “On kinematic waves II: a theory of traffic flow on long crowded roads.,” *Proceedings of The Royal Society A: Mathematical, Physical and Engineering Sciences,* vol. 229, pp. 317-345, 10 May 1955. |
| [2] | P. I. Richards, “Shockwaves on the Highway,” *Operations Research,* vol. 4, no. 1, pp. 42 - 51, 1 Febuary 1956. |
| [3] | W. Daamen, P. H. L. Bovy and S. P. Hoogendoorn, “Modelling Pedestrians in Transfer Stations,” in *Pedestrian and Evacuation Dynamics*, M. Schreckenberg and S. D. Sharma, Eds., Duisburg, Springer Verlag, 2003, pp. 59 - 74. |
| [4] | A. Schradschneider, H. Klupfel, T. Kretz, C. Rogsch and A. Seyfried, “Fundamentals of pedestrian and evacuation dynamics,” in *Multi Agent Systems fo Traffic and Transportation Engineering*, 2009, pp. 124-154. |
| [5] | N. Pelechano and A. Malkawi, “Evacuation Simulation models: Challenges in modelling high rise building evacuation with cellular automata approaches,” *Automation in Construction,* vol. 17, no. 4, pp. 377-385, May 2008. |
| [6] | UK Government, “Electricity generation and supply figures for Scotland, Wales, Northern Ireland and England, 2010 to 2013,” UK Governement, 2014. |
| [7] | SMMT, “June 2015 - EV registrations,” 2015. [Online]. Available: http://www.smmt.co.uk/2015/07/june-2015-ev-registrations/. [Accessed 25 July 2015]. |
| [8] | Science News, “Ancient Rome forbade downtown traffic in day,” *Science News,* vol. 78, no. 5, 30 July 1960. |
| [9] | D. Metz, The Limits to Travel: How Far Will You Go?, Taylor and Francis, 2012. |
| [10] | s. tdes, “No Traffic Lights Vs Traffic Lights,” 2010. |
| [11] | DVSA (Driver and Vehicle Standards Agency), “Guide to lorry types and weights,” DfT, London, 2013. |
| [12] | K. Pandit, D. Ghosal, H. M. Zhang and C.-N. Chuah, “Adaptive Traffic Signal Control With Vehicular Ad hoc Networks,” *IEEE Transactions on Vehicular Technology,* vol. 62, no. 4, pp. 1459-1471, May 2013. |
| [13] | L. Singh, S. Tripathi and H. Arora, “Time Optimization for Traffic Signal Control Using Genetic Algorithm,” *International Journal of Recent Trends in Engineering,* vol. 2, no. 2, pp. 4-6, November 2009. |
| [14] | S. Sayyed, P. Date, R. Gautam and G. Bhandari, “Design of Dynamic Traffic Signal Control System,” *International Journal of Engineering Research & Technology,* vol. 3, no. 1, pp. 933-938, January 2014. |
| [15] | J. Gregoire, X. Qian, E. Frazzoli, A. de La Fortelle and T. Wongpiromsarn, “Capacity-aware back-pressure traffic signal control,” *IEEE Transactions on Control of Network Systems,* no. 99, 08 December 2014. |
| [16] | L. Qi and M. Zhou, “Modeling and Control of Urban Road Intersections with Incidents via Deterministic and Stochastic Petri Nets,” in *12th IEEE International Conference on Networking, Sensing and Control*, Taipei, 2015. |
| [17] | D. Manolis, T. Pappa, C. Diakaki, I. Papamichail and M. Papageorgiou, “Centralised vs decentralised signal control of large-scale urban road networks in real time: A simulation study,” *IET Intelligent Transport Systems,* vol. 12, no. 8, pp. 891-900, 2018. |
| [18] | R. Parmar, B. Trivedi and A. Stevanovic, “ A Model with Traffic Routers, Dynamically Managing Signal Phases to Address Traffic Congestion in Real Time,” *Journal of Transportation Technologies,* vol. 8, pp. 75-90, 2018. |
| [19] | J. Hardy and L. Liu, “Reducing Vehicular Traffic Congestion Using Available Forward Road Capacity Detection,” in *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, Liverpool, 2015. |
| [20] | J. Hardy and L. Liu, “Available Forward Road Capacity Detection Algorithms to Reduce Urban Traffic Congestion,” in *10th International Conference on Cyber, Physical and Social Computing - CPSCom2017*, Exeter, 2017. |
| [21] | DfT, “An introduction to the Department for Transport’s road congestion statistics,” Department for Transport, London, 2013. |
| [22] | M. Cassini, “Experiment Town in England turns off traffic lights surprising results,” 2009. |
| [23] | L. H. o. Richmond, “Accidents by Design: The Holmes Report on 'shared space' in the United Kingdom,” 2015. |
| [24] | Poynton Town Council, “Minutes of the meeting of the Community, Order and Public Safety Committee held at 8pm on Monday 3rd September 2012 in the Council Chamber, Civic Hall, Poynton,” Poynton, 2012. |
| [25] | M. Cassini, “Poynton Regenerated,” 2011. |
| [26] | A. Hamilton, B. Waterson, T. Cherrett, A. Robinson and I. Snell, “The evolution of urban traffic control: changing policy and technology,” *Transportation Planning and Technology,* vol. 36, no. 1, pp. 24 - 43, 2013. |
| [27] | MIZAR Automazione S.p.A., “UTOPIA – Urban Traffic Control System Architecture,” MIZAR, Verona, 2012. |
| [28] | Vehicle Detector Clearinghouse, “A Summary of Vehicle Detection and Surveillance Technologies used in Intelligent Transportation Systems,” New Mexico State University, Las Cruces , 2007. |
| [29] | Sanderson Associates, “Microprocessor Optimised Vehicle Actuation,” 2016. [Online]. Available: http://www.traffic-signal-design.com/microprocessor\_optimised\_vehicle\_actuation\_mova.htm. [Accessed 18 August 2016]. |
| [30] | HA, “Installation Guide for MOVA,” Highways Agency, London, 2005. |
| [31] | M. R. Crabtree, “MOVA Traffic Control Manual,” TRL Ltd, Wokingham, 2011. |
| [32] | Department for Transport, “General Principles of Traffic Control by Traffic Light Signals Part 1 of 4,” Department for Transport, London, 2006. |
| [33] | SCATS, “The New Generation SCATS 6 Functional Description”. |
| [34] | D. Robertson and R. Bretherton, “Optimizing networks of traffic signals in real time-the SCOOT method,” *IEEE Transactions on Vehicular Technology,* vol. 40, no. 1, pp. 11-15, Feb 1991. |
| [35] | Department for Transport, “Traffic Advisory Leaflet 7/99 - The "SCOOT" Urban Traffic Control System,” Department for Transport, London, 1999. |
| [36] | TRL Software, “SCOOT,” 2015. [Online]. Available: https://trlsoftware.co.uk/products/traffic\_control/scoot. [Accessed 10 February 2016]. |
| [37] | P. T. Martin, “Scoot, an overview,” University of Utah, Utah, 2001. |
| [38] | SCOOT, “How SCOOT Works,” 2016. [Online]. Available: http://www.scoot-utc.com/DetailedHowSCOOTWorks.php. [Accessed 21 Sept 2016]. |
| [39] | DfT, “The SCOOT Urban Traffic Control System,” Department for Transport, London, 1995. |
| [40] | Roads and Maritime Services, NSW Australia, SCATS 6 functional description, Canberra: Roads and Maritime Services, NSW Australia, 2013, p. 33. |
| [41] | S. Bayless and A. Guan, “Connected Vehicle Technical Insights, Vehicle Applications and Wireless Interoperability,” ITS-America, 2015. |
| [42] | M. Chitra, S. Sivasathya and B. Muthamizh, “Performance Analysis of Topology based Routing Protocols in VANET,” *International Journal of Computer Science and Mobile Computing,* vol. 3, no. 3, pp. 966-975, March 2014. |
| [43] | F. J. Martinez, C. K. Toh, J. C. Cano, C. T. Calafate and P. Manzoni, “A survey and comparative study of simulators for vehicular ad hoc networks (VANETs),” *Wireless Communications and Mobile Computing,* vol. 11, no. 7, pp. 813-828, July 2011. |
| [44] | R. B. Noland, “Relationships between highway capacity and induced vehicle travel,” *Transportation Research Part A 3,* vol. 35, pp. 47-72, 2001. |
| [45] | D. Braess, “Über ein Paradoxon aus der Verkehrsplanung,” *Unternehmensforschung,* vol. 12, pp. 258-268, 1968. |
| [46] | D. Braess, A. Nagurney and T. Wakolbinger, “On a Paradox of Traffic Planning,” *Transportation Science,* vol. 39, no. 4, pp. 446-450, 2005. |
| [47] | E. N. Lorenz, “Deterministic Nonperiodic Flow,” *Journal of the Atmospheric Sciences,* vol. 20, pp. 130-141, 1963. |
| [48] | A. Adewumi, J. Kagamba and A. Alochukwu, “Application of Chaos Theory in the Prediction of Motorised Traffic Flows on Urban Networks,” *Mathematical Problems in Engineering,* vol. 2016, no. 78, pp. 1-15, 2015. |
| [49] | A. Johansen and D. Sornette, “Stock market crashes are outliers,” *The European Physics Journal B,* vol. 1, no. 2, pp. 141-143, 1998. |
| [50] | C. Frazier and K. K. M, “Chaos Theory and Transportation Systems: An Instructive Example,” *Transportation Research Record,* vol. 1897, pp. 9-17, 2004. |
| [51] | A. K. Erlang, “The Theory of Probabilities and Telephone Conversations,” *Nyt Tidsskrift for Matematik,* vol. 20, no. 8, pp. 33-39, 1909. |
| [52] | A. Kouvelas, D. Triantafyllos and N. Geroliminis, “Hierarchical Control Design for Large-scal Urban Road Traffic Networks,” in *Transportation Research Board 97th Annual Meeting*, Washington, D.C, 2018. |
| [53] | C. A. Petri, “Kommunikation mit Automaten,” 1962. |
| [54] | C. A. Petri, “Communication with Automata - Technical Report RADC-TR-65-377 vol 1, supl 1,” New York, 1966. |
| [55] | C. A. Petri, “Fundamentals of a Theory of Asynchronous Information Flow,” in *IFIP Congress 62, Munchen*, Amsterdam, 1963. |
| [56] | J. Škrinjar, B. Abramović and N. Brnjac, “The use of game theory in urban transport planning,” *Tehnicki Vjesnik,* vol. 22, no. 6, pp. 1617-1621, 2015. |
| [57] | R. Hernández, C. Cárdenas and D. Muñoz, “Game theory applied to transportation systems in Smart Cities: analysis of evolutionary stable strategies in a generic car pooling system,” *International Journal on Interactive Design and Manufacturing (IJIDeM),* vol. 12, no. 1, pp. 179-185, 2018. |
| [58] | D. W. Bromley, R. Buehler, T. Godefrooij, M. Kirkels, N. Quental, K. Kusakabe, A. Vasishth, J. D. Tejada, B. Geys, M. D. Leiren, M. S. Khan, E. Mwendera, S. Melia, M. Brussel, M. Zuidgeest, F. De Souza, S. R. Acharya, P. Dirgahayani and M. Martiskainen, “Viewpoints: How can the stigma of public transport as the ‘poor man's vehicle’ be overcome to enhance sustainability and climate change mitigation?,” vol. 34, no. 4, pp. 327-331, November 2010. |
| [59] | J. Casas, J. L. Ferrer, D. Garcia, J. Perarnau and A. Torday, “Traffic Simulation with Aimsun,” in *Fundamentals of Traffic Simulation*, New York, Springer, 2010, pp. 173-232. |
| [60] | Department for Transport, “Road reported casualties Great Britain 2012,” Office of National Statistics, London, 2013. |
| [61] | Society of Motor Manufacturers and Traders, “motor industry facts 2013,” the society of motor manufacturers and traders Limited, London, 2013. |
| [62] | Society of Motor Manufacturers and Traders , “Motor Industry Facts 2016,” The Society of Motor Manufacturers and Traders Limited, 2016. |
| [63] | T. Barlow, “Escape the cost and lost time of traffic congestion,” 22 October 2011. [Online]. Available: http://www.forbes.com/sites/tombarlow/2011/10/22/escape-the-cost-and-lost-time-of-traffic-congestion/. [Accessed 5 July 2015]. |
| [64] | D. Shrank, W. Eisele and T. Lomax, “TTI’s 2012 Urban Mobility Report,” Texas A&M Transportation Institute, Texas, 2012. |
| [65] | USA SSA, “Average Wage Index,” 2013. [Online]. Available: http://www.ssa.gov/oact/cola/awidevelop.html. |
| [66] | USA Department of Energy, “Fact #835: August 25, Average Historical Annual Gasoline Pump Price, 1929-2013,” 2015. [Online]. Available: http://energy.gov/eere/vehicles/fact-835-august-25-average-historical-annual-gasoline-pump-price-1929-2013. |
| [67] | Inrix, “INRIX Research,” Inrix, [Online]. Available: http://inrix.com/research/. |
| [68] | M. M. Minderhoud, H. Botma and P. H. L. Bovy, “Assessment of Roadway Capacity Estimation Methods,” *Transportation Research Record: Journal of the Transportation Research Board,* vol. 1572, pp. 59-67, 1997. |
| [69] | J. van Rijn, “publications,” 2004. [Online]. Available: http://www.indevelopment.nl/PDFfiles/CapacityOfRroads.pdf. [Accessed 27 Oct 2014]. |
| [70] | TRB Publications, Highway Capacity Manual 2010, 5th ed., 2010. |
| [71] | TRB Publications, Highway Capacity Manual 2000, TRB Publications, 2000. |
| [72] | D.-H. Ha, M. Aron and S. Cohen, “Time headway variable and probabilistic modeling,” *Transportation Research Part C: Emerging Technologies,* vol. 25, pp. 181-201, Dec 2012. |
| [73] | S. Oh and H. Yeo, “A Study on the Relationship Between Capacity Drop and the Number of Lanes in Freeway Merging Sections,” in *Eastern Asia Society for Transportation Studies*, 2011. |
| [74] | S. Koole, “How We Came to Mind the Gap: Time, Tactility, and the Tube,” *Twentieth Century British History,* vol. 27, no. 4, p. 524–554, Dec 2016. |
| [75] | S. M. Abtahi, M. Tamannaei and H. Haghshenash, “Analysis and modeling time headway distributions under heavy traffic flow conditions in the urban highways: case of Isfahan,” *Transport,* vol. 26, no. 4, pp. 375-382, 2012. |
| [76] | S. Moridpour, “Evaluating the Time Headway Distributions in Congested Highways,” *Journal of Traffic and Logistics Engineering,* vol. 2, no. 3, pp. 224-229, 2014. |
| [77] | J. I. Lake, K. S. LaBar and W. H. Meck, “Emotional modulation of interval timing and time perception,” *Neuroscience and Biobehavioural Reviews,* vol. 64, pp. 403-420, 2016. |
| [78] | S. Anand, S. V. C. Sekhar and K. M. R, “development of passenger car unit (pcu) values for malaysia,” *Joumal of the Eastern Asia Society for Transportation Studies,* vol. 3, no. 3, pp. 73 - 80, Sept 1999. |
| [79] | Transport for London, “Traffic Modelling Guidlines,” London, 2010. |
| [80] | S. Marlina, “Understanding the dynamics of truck traffic on freeways by evaluating truck passenger car equivalent (pce) in the highway capacity manual (hcm),” University of Colorado, Colorado, 2010. |
| [81] | J. Kim and L. Elefteriadou, “Estimation of Capacity of Two-Lane Two-Way Highways Using Simulation Model,” *Journal of Transportation Engineering,* vol. 136, no. 1, pp. 61-66, January 2010. |
| [82] | X. Yang and N. Zhang, “The marginal decrease of lane capacity with the number of lanes on highway,” in *Proceedings of the Eastern Asia Society for Transportation Studies*, 2005. |
| [83] | R. M. Brooks, “Influence of vehicular action on PCU values,” in *The Institution of Civil Engineers Transport 163*, 2010. |
| [84] | TfL, “Traffic Modelling Guidelines TfL Traffic Manager and Network Performance Best Practice,” Transport for London, London, 2010. |
| [85] | A. Srivastava and N. Geroliminis, “Empirical observations of capacity drop in freeway merges with ramp control and integration in a first-order model,” *Transportation Research Part C,* vol. 30, pp. 161-177, 2013. |
| [86] | Department for Transport, The Official Highway Code, 2015 ed., London: HMSO, 2015. |
| [87] | A. Mehmood and S. M. Easa, “Modeling Reaction Time in Car-Following Behaviour Based on Human Factors,” *World Academy of Science, Engineering and Technology,* vol. 3, pp. 651 - 659, Sept 2009. |
| [88] | R. K. Kamalanathsharma, “Acceleration and Deceleration Characteristics of Vehicles at Intersections – A Literature Review,” c2012. |
| [89] | T. J. Gates, D. A. Noyce and L. Laracuente, “Analysis of Dilemma Zone Driver Behavior at Signalized Intersections,” *Transportation Research Record: Journal of the Transportation Research Board,* vol. 2030, pp. 29-39, 2007. |
| [90] | B. Greenshields, “A Study of Traffic Capacity,” in *14th Annual Meeting of the Highway Research Board*, 1935. |
| [91] | S. Smulders, “Control of freeway traffic flow by variable speed signs,” *Transportation Research Part B,* vol. 24, no. 2, p. 111–132, 1990. |
| [92] | J. del Castillo, “Three new models for the flow–density relationship: derivation and testing for,” *Transportmetrica,* vol. 8, no. 6, p. 443–465, 2012. |
| [93] | D. Helbing, “Derivation of a fundamental diagram for urban traffic flow,” *The European Physical Journal B,* vol. 70, no. 2, pp. 229-241, 2009. |
| [94] | F. van Wageningen-Kessels, H. Van Lint, K. Vuik and S. Hoogendoorn, “Genealogy of Traffic Flow Models,” *EURO Journal on Transportation and Logistics,* vol. 4, no. 4, pp. 445-473, 2015. |
| [95] | Highways Agency, “TR2500 A - Specification for Traffic Signal Controller,” Highways Agency, London, 2005. |
| [96] | UK Government, “The Traffic Signs Regulations and General Directions 2002,” Crown, London, 2002. |
| [97] | Department for Transport, “General Principles of Traffic Control by Traffic Light Signals Part 2 of 4,” Department for Transport, London, 2006. |
| [98] | D. Herman, “Introducing short-term brands: A new branding tool for a new consumer reality,” *Journal of Brand Management,* vol. 7, no. 5, pp. 330-340, 2000. |
| [99] | Inrix, “Press Releases,” Inrix, 14 october 2014. [Online]. Available: Traffic Congestion to Cost the UK Economy More Than £300 Billion Over the Next 16 Years. [Accessed Mar 2018]. |
| [100] | T. Litman, “Generated traffic and induced travel,” *ITE Journal,* vol. 71, no. 4, pp. 38-47, April 2001. |
| [101] | H. Liu, Y. Zhang, Q. Zhu and J. Chu, “Environmental efficiency of land transportation in China: A parallel slack-based measure for regional and temporal analysis,” *Journal of Cleaner Production,* vol. 142, pp. 867-876, 2017. |
| [102] | “American Council for an Energy Efficient Economy,” 2018. [Online]. Available: https://aceee.org/topics/transportation-system-efficiency. [Accessed 2 Oct 2018]. |
| [103] | P. W. Cooper, “Further Commentary on “Through the Earth in Forty Minutes”,” *American Journal of Physics,* vol. 34, no. 8, p. 703, 2005. |
| [104] | Y. E. Kozlovsky, The Superdeep Well of the Kola Peninsula, Berlin: Springer-Verlag, 1987. |
| [105] | W. Mooney, G. Laske, G. Masters and K. H. Favret, “How Thick is the Earth's Crust?,” *Journal of Geophysical Research Atmospheres,* vol. 103, pp. 172-181, 1998. |
| [106] | M. Dubarry, V. Svoboda, R. Hwu and B. Y. Liaw, “Capacity loss in rechareable lithium cells during life cycle testing: The importance of determining state-of-charge,” *Jornal of Power Source,* vol. 174, pp. 1121-1125, 2007. |
| [107] | K. Fisher, E. Wallen, P. P. Laenen and C. M, “Battery Waste Management Life Cycle Assessment,” Defra, London, 2006. |
| [108] | K. Aguirre, L. Eisenhardt, C. Lim, B. Nelson, A. Norring, P. Slowik and N. Tu, “Lifecycle analysis comparison of a battery electric vehicle and a conventional gasoline vehicle,” California Air Resource Board, California, 2012. |
| [109] | A. F. Sherwani, J. A. Usmani and Varun, “Life cycle assessment of solar PV based electricity generation systems: A review,” *Renewable and Sustainable Energy Reviews,* vol. 14, no. 1, pp. 540-544, January 2010. |
| [110] | D. Qu, M. Wan, Z. Wang, X. Xu and J. Wang, “Green Wave Coordinate Control Method for Arterial Traffic Based on Traffic Wave Theory,” *Journal of Highway and Transportation Research and Development,* vol. 11, no. 4, pp. 65-73, 2017. |
| [111] | X. Kong, G. Shen, F. Xia and C. Lin, “Urban Arterial Traffic Two-direction Green Wave Intelligent Coordination Control Technique and Its Application,” *International Journal of Control, Automation, and Systems,* vol. 9, no. 1, pp. 60-68, 2011. |
| [112] | X. P. Wu, S. Deng, X. H. Du and J. Ma, “Green-Wave Traffic Theory Optimization and Analysis,” *World Journal of Engineering and Technology,* vol. 2, pp. 14-19, 2014. |
| [113] | M. Bazan, P. Ciskowski, K. Halawa, T. Janiczek, Ł. Madej and A. Rusiecki, “Green wave optimisation,” *Transport System Telematics,* vol. 9, no. 3, pp. 3-8, 2016. |
| [114] | O. Tomescu, V. Tomás, F. Nemtanu and I. Buațroș, “Traffic Signal Green-Wave Control Strategy Based on Divers’ Behaviors,” *Analele Universituactii Eftimie Murgu Recsicta. Fascicula de Inginerie.,* vol. XVIII, pp. 146-156, 2011. |
| [115] | B. S. Kerner, Breakdown in Traffic Networks, Berlin: Springer, 2017. |
| [116] | B. S. Kerner, Introduction to modern traffic flow theory and control, Berlin: Springer, 2009. |
| [117] | F. Castillo, B. A. Toledo, V. Muñoz, J. Rogan, R. Zarama, J. Penagos, M. Kiwi and J. A. Valdivia, “Spatiotemporal complexity of a City Traffic Jam,” *Journal of Cellular Automata,* vol. 0, pp. 1-18, 2015. |
| [118] | P. C. Blumenfeld, E. Soloway, R. W. Marx, J. S. Krajcik, M. Guzdial and A. Palincsar, “Motivating Project Based Learning: Sustaining the Doing, Supporting the Learning,” *Eductaional Psychologist,* vol. 26, no. 3 & 4, pp. 369-398, 1991. |
| [119] | google, “traffic overlay derby,” 18 January 2019. [Online]. Available: https://www.google.co.uk/maps/@52.9265644,-1.4781963,17.39z/data=!5m1!1e1. [Accessed 18 Jan 2019]. |
| [120] | Microsoft Bing, “Traffic Derby,” 18 January 2019. [Online]. Available: https://www.bing.com/maps/traffic. [Accessed 18 jan 2019]. |
| [121] | D. Krajzewicz, J. Erdmann, M. Behrisch and L. Bieker, “Recent Development and Applications of SUMO - Simulation of Urban MObility,” *International Journal On Advances in Systems and Measurements,* vol. 5 (3&4), pp. 128-138, December 2012. |
| [122] | F. Kröger, “Automated Driving in Its Social, Historical and Cultural Contexts,” in *Autonomous Driving: Technical, Legal and Social Aspects*, Berlin, Heidelberg, Springer Berlin Heidelberg, 2016, pp. 41 - 68. |
| [123] | National Transportation Safety Board, “ ​Preliminary Report Highway: HWY18MH010,” National Transportation Safety Board, 2018. |
| [124] | IHS Global Insight, “Assessment of the Effectiveness of Scrapping Schemes for Vehicles. Economic, Environmental, and Safety Impacts,” 2010. |
| [125] | European Conference of Ministers of Transport, “CLEANER CARS Fleet Renewal and Scrappage Schemes,” Paris, 1999. |
| [126] | IHS Global Insight, “Assessment of the Effectiveness of Scrapping Schemes for Vehicles. Country Profile Annex,” 2010. |
| [127] | J. Xu, D. Deng, U. Demiryurek, C. Shahabi and M. van der Schaar, “Mining the Situation: Spatiotemporal Traffic Prediction With Big Data,” *IEEE Journal of Selected Topics in Signal Processing,* vol. 9, no. 4, pp. 702-715, 2015. |
| [128] | Department for Transport, “Reported Road Casualties Great Britain: 2014 Annual Report,” September, London, 2015. |
| [129] | Department for Transport, “Road Traffic Estimates: Great Britain 2014 (R),” DfT, London, 2015. |
| [130] | DfT, “Casualties involved in reported road accidents (RAS30),” 30 June 2016. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment\_data/file/461911/ras30019.xls. [Accessed 2 Sept 2016]. |
| [131] | Society of Automotive Engineers, “Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems,” 2014. |
| [132] | The Guardian, “Number of planes to double in next two decades, Boeing forecasts,” 2013. [Online]. Available: https://www.theguardian.com/business/2013/jun/11/boeing-commercial-planes-double-asia-pacific. [Accessed 20 July 2016]. |
| [133] | Department for Transport, “Vehicle Licencing Statistics: 2013,” Department for Transport, London, 2014. |
| [134] | A. Mahmoudabadi and S. Andalibi, “The Assessment of Applying Chaos Theory for Daily Traffic Estimation,” in *Proceedings of the 2014 International Conference on Industrial Engineering and Operations Management*, Bali, Indonesia, 2014. |
| [135] | L. A. Safonov, E. Tomer, V. V. Strygin, Y. Ashkenazy and S. Havlin, “Multifractal chaotic attractors in a system of delay-differential equations,” *Chaos,* vol. 12, no. 4, pp. 1006-1014, 2002. |
| [136] | HA, “Traffic Control System Design for All Purpose Roads (Compendium of Examples),” Highways Agency, London, 2003. |
| [137] | M. Metkari, A. K. Budcar and A. K. Maurya, “Review of Passenger Car Equivalence Studies in Indian Context,” *IJCA Proceedings on International Conference on Emerging Frontiers in Technology for Rural Area,* vol. EFITRA, pp. 19-23, May 2012. |
| [138] | R. Bellman, “On A routing Problem,” *Quarterly of Applied Mathematics,* vol. 16, no. 1, pp. 87-90, 1958. |
| [139] | L. R. Ford, “Network Flow Theory - Paper P-923,” The Rand Corporation, Santa Monica, 1956. |
| [140] | E. W. Dijkstra, “A Note on Two Problems in Connexion with Graphs,” *Numerische Mathematik,* vol. 1, pp. 269 - 271, 1959. |
| [141] | UN, “Sustainable Transport,” United Nations, June 1992. [Online]. Available: http://sustainabledevelopment.un.org/index.php?menu=238. [Accessed 1 September 2014]. |
| [142] | UN, “United Nations Conference on Environment & Development Rio de Janerio, Brazil, 3 to 14 June 1992 AGENDA 21,” Rio de Janerio, 1992. |
| [143] | DfT, “National Travel Survey 2012,” London, 2013. |
| [144] | DfT, “National Travel Survey 2013,” London, 2014. |
| [145] | DfT, “Commuting and bussiness travel factsheet - April 2011,” London, 2011. |
| [146] | UN, “Back to Our Common Future. Sustainable Development in the 21st century (SD21) project. Summary for policymakers,” New York, 2012. |
| [147] | M. McCarthy, “The benefit of seat belt legislation in the United Kingdom,” *Journal of Epidemiology and Community Health,* vol. 43, no. 3, pp. 218-222, 1989. |
| [148] | ONS - Office for National Statistics, “Area:Cambridge(Local Authority); Economic Activity 2011,” March 2011. [Online]. Available: http://www.neighbourhood.statistics.gov.uk/dissemination/LeadTableView.do?a=5&b=6275044&c=cambridge&d=13&e=61&g=6405296&i=1001x1003x1032x1004&m=0&r=1&s=1409323508439&enc=1&dsFamilyId=2521. [Accessed 28 Aug 2014]. |
| [149] | University of Cambridge, “Motor Proctor :: University of Cambridge,” November 2013. [Online]. Available: http://www.admin.cam.ac.uk/offices/proctors/motor/. [Accessed 28 Aug 2014]. |
| [150] | ONS - Office for National Statistics, “Area Cambridge;Car or Van Availability, 2011,” Mar 2011. [Online]. Available: http://www.neighbourhood.statistics.gov.uk/dissemination/LeadTableView.do?a=5&b=6275044&c=cambridge&d=13&e=61&g=6405296&i=1001x1003x1032x1004&m=0&r=1&s=1409323508439&enc=1&dsFamilyId=2511. [Accessed 28 Aug 2014]. |
| [151] | Thatcham Research , “What is Autonomous Emergency Braking (AEB)?,” 2016. [Online]. Available: http://www.thatcham.org/car-safety/automatic-braking/. [Accessed 14 October 2016]. |
| [152] | Ford Motor Company Ltd, “New Ford Fiesta ‘Active City Stop’ safety tech captures and processes 15 images in the blink of an eye,” 2016. [Online]. Available: http://www.ford.co.uk/FordFleet/NewsAndReviews/FordForBusiness/2012/November/FiestaActiveCityStop. [Accessed 14 October 2016]. |
| [153] | Euro NCAP, “FIAT City Brake Control,” Eur NCAP, 2013. |
| [154] | M. Green, “'How Long Does It Take To Stop?' Methodological Analysis of Driver Perception-Brake Times,” *Transportation Human Factors,* vol. 2, no. 3, pp. 195-216, 2000. |
| [155] | K. Bevrani, E. Chung and M. Miska, “Evaluation Of The GHR Car Following Model For Traffic Safety Studies,” in *25th ARRB Conference – Shaping the future: Linking policy, research and outcomes*, Perth, Australia, 2012. |
| [156] | Federal Highways Agency, “Transims background,” 2015. [Online]. Available: http://www.fhwa.dot.gov/planning/tmip/transims/background.cfm. [Accessed 29 July 2016]. |

1. Traffic Light Control Algorithms

Timed Algorithm

This is the simplest traffic light control algorithm. Each light phase is sequential and active (showing green) until the timer expires. In a real environment, the timing of each phase of each junction would be determined in advance based on actual vehicle counts undertaken at the roadside. The timing would be based on day, time and ingress and egress routes. From the collected data, flows and densities are calculated and used to create timing schemes for “normal” conditions. The schemes are specific to a particular junction at a particular time.

From a simulation viewpoint, fixed timing can be problematic. To be effective, each junction should have an independent timing scheme and there will either be no synchronisation or absolute offset synchronisation between adjacent controllers. The simulation fails in many of these respects; every controller has the same sequence, same timing and is synchronised. While it would be possible to create individual plans for the simulation, this would be time consuming and there would be little benefit to the overall research. The synchronisation problem would be more difficult to resolve as the simulation cycle time would prevent timing drift which could be present in a real system. Timing drift could be the result of inaccurate clock rates, differences in clock cycles required for each phase or software aging due minor calculation, numeric representations and storage deviations (reference to software aging here) Items contained in curly brackets {} are specific to the simulation. The statement “force change” means that the current phase should advance to the next phase.

Algorithm:

{After Aimsun loads the simulation}

define normal phase parameters of tl

{before Aimsun starts the simulation}

initialise simulation result file

{at the END of every simulation cycle}

for each junction:

Clear all status flags

Save pre-calculation data to file

Calculate required changes.

For current phase:

if light has been active for more than absolute max time, force change

Save post-calculation data to file

Modified VA Algorithm

The following is the modified vehicle actuated (VA) algorithm implemented for some of the simulation testing. The control program constructed from the algorithm is used in a simulation however it can equally be used in a real controller. In the simulation all traffic lights use the same control sequence and same algorithm. To reduce coding and simplify changes, each junction is tested in sequence at the end of each simulation cycle and changes are made prior to the start of the next cycle.

The modified VA algorithm is based on the action of a real VA control. This version is modified to prioritise opposing flow queue detection over continuity of flow in the current direction. The system requires two maximum timers. The first timer prevents continual short duration signal activations but forces a pre-emptive change to the next sequence if there is no current traffic flow in the active direction. The second timer sets the absolute maximum time that a green is shown in any direction. In a real environment the timers would need to be configured to the most suitable values using similar methods to timed sequence.

The AFRC algorithm is built over this controller or the traditional VA controller. It could also be built over a Microprocessor Optimised VA (MOVA) controller and provide similar benefits. The improvement offered by AFRC should be applicable to any traffic controller which uses approach data as the prime (or only) sequencing input.

Items contained in curly brackets {} are specific to the simulation

Algorithm:

{After Aimsun loads the simulation}

define normal phase parameters of tl

{before Aimsun starts the simulation}

initialise simulation result file

create state matrix (20 data for each junction)

{at the END of every simulation cycle}

for each junction:

get junction information from lookup table

update state matrix with latest values

for each approach:

populate matrix with detector value

Clear all status flags

Save pre-calculation data to file

Calculate required changes:

For current phase:

if light has been active for more than absolute max time, force change

if light has been active for more than the minimum time

if vehicle waiting from other approach

if light has been active for longer than normal period, force change

if nothing has passed from this approach, force change

Save post-calculation data to file

Traditional VA Algorithm

The following is an implementation of the traditional vehicle actuated (VA) algorithm. The control program constructed from the algorithm is used in simulation however it can equally be used in a real controller. In the simulation all traffic lights use the same control sequence and same algorithm. To reduce coding and simplify changes in the simulation, each junction is tested in sequence at the end of each simulation cycle and changes are made prior to the start of the next simulation cycle. The VA algorithm is based on the action of a real control. Real VA systems maintain a green while traffic is flowing and test for queues at other approaches only when traffic flow has stopped flowing. A signal change is made either when there is no traffic flow and an opposing queue or when the absolute maximum timer expires, the timers may be different in opposing directions. The sequence of phases is fixed. A junction is determined to be congested if traffic is still flowing when the absolute maximum time has expired. Under busy or congested flows the action of a VA system will tend to run to the maximum timer. Within the simulation, and potentially in real systems, a vehicle which is stationary over a loop detector will register a presence. The action of the detector makes a presence indistinguishable from a flow and therefore a stationary queue at a green signal can, depending on precise vehicle position, extend the green to the maximum possible time with no benefit but with a detrimental impact to the opposing flow.

The AFRC algorithm is built over this controller for some of the theoretical understanding and simulation experiments. Inclusion of two VA systems is intended to show that for any system deriving sequencing from approach sensing, the base controller system is irrelevant in obtaining the benefit of AFRC detection.

Algorithm:

{After Aimsun loads the simulation}

define normal phase parameters of tl

{before Aimsun starts the simulation}

initialise simulation result file create state matrix (20 data for each junction)

{at the END of every simulation cycle}

for each junction:

get junction information from lookup tableupdate state matrix with latest values

for each approach

populate matrix with detector value

Clear all status flags

Save pre-calculation data to file

Calculate required changes:

for current phase:

if light has been active for more than absolute max time, force change

if light has been active for more than the minimum time

if vehicle waiting from other approach

if light has been active for longer than normal period, force change

if nothing has passed from this approach, force change

Save post-calculation data to file

AFRC Algorithm

The following is the full AFRC algorithm. The control program constructed from the algorithm is used in a simulation however it can equally be used in a real controller. In the simulation all traffic lights use the same control sequence and same algorithm. To reduce coding and simplify changes in the simulation, each junction is tested in sequence at the end of each simulation cycle and changes are made prior to the start of the next cycle.

The AFRC algorithm is an extension to the modified VA algorithm which aims to remove deficiencies of the VA system and improve the scheduling efficiency. While this algorithm is based on VA, it can easily be adapted to MOVA which has several of the same inherent deficiencies.

The cycle time for the AFRC system is far less critical to the extent that it is largely irrelevant if the detectors are consistent, repeatable and reliable. The AFRC system is self-tuning under all situations as it is reactive to actual current conditions and events and is not prediction based. The AFRC system is intended to reduce lost time due to unrealistic signals, reduce lost time due to signal transitions and increase the effective utilisation of the contended junction space. Sequencing and timing in response to unfeasible junction movements will not increase driver anxiety and cannot give rise to increased civil disobedience. The signal action is to reinforce a physical inability to progress through a junction rather than to enforce a restriction on an otherwise feasible operation. The system is also intended to be implemented with minimal prior knowledge for a junction.

Items contained in curly brackets {} are specific to the simulation

Algorithm:

{After Aimsun loads the simulation}

define normal phase parameters of tl

{before Aimsun starts the simulation}

initialise simulation result file

create state matrix (20 data for each junction)

{at the END of every simulation cycle}

for each junction:

get junction information from lookup table

update state matrix with latest values

For each exit:

If blocked, set “upstream request” matrix flag for next junction downstream

For each approach:

Populate matrix with detector value

Clear all status flags

Save pre-calculation data to file

Create blocked exit data for logging

Set all phases as feasible at start of calculation

Set phases with blocked major exits as not feasible

Set phases that cannot clear upstream request as not feasible

Read neighbour request from all neighbours

Calculate required changes. For current phase:

if phase 1 – 4 and phase clears upstream flow, clear upstream request

if light has been active for more than absolute max time, force change

if this phase main exit is blocked, force change

if light has been active for more than the minimum time

if next upstream junction is requesting service, force change

if vehicle waiting from other approach

if light has been active for longer than normal period, force change

if nothing has passed from this approach, force change

Save post-calculation data to file

If change required

Determine first sequential phase feasible

Change to phase

Reset phase start time

If no feasible phase

If phase changed at previous request

Repeat current state

If current state has already repeated

Scavenge for clear exits, prefer phases 1-4

Only select phases with approach queue

Last resort : Repeat current state (or all-red unpowered traffic)

1. Traffic signal terminology

The following is selection of definitions taken from Sanderson Associates website, <http://www.traffic-signal-design.com/terminology_main.htm>. The definitions are used as clarification of the rest of their website, but do serve as an adequate ready reference. All pictures and diagrams have been removed.

|  |  |
| --- | --- |
| Phase | An indication shown to a particular traffic or pedestrian link. Each phase at a junction exists as an electrical circuit from the controller and feeds one or more signal heads. A phase can apply to a single aspect (filter or indicative arrow), a two aspect head (pedestrians) or a three aspect head (red / amber / green where the green could be a directional arrow). |
| Stage | A series of non-conflicting phases which run together. A stage starts when the last of the phases commences and then ends when the first of the phases terminates. |
| Stage Diagram | A diagram using arrows which shows the phase movements permitted in each stage. |
| Cycle | One complete sequence of the operation of traffic signals. |
| Cycle Time | The time taken to complete one cycle. In the UK the maximum cycle time is usually 120 seconds, but this is often lowered to a maximum 90 seconds where pedestrian facilities are present. Signal controlled roundabouts often have shorter cycle times as they often have limited stacking space on the circulatory carriageway which quickly fills up and blocks the exit from the roundabout. |
| All Red | A condition of traffic signals where all traffic movements are shown a red signal. This is often to allow a pedestrian stage to run. |
| All Red Period | Period during the change from one phase green to the next phase green where all phases show red. |
| Intergreen Period | The time between end of right of way for a phase and the start of the right of way for the next phase. These comprise of the 3s leaving amber for the phase losing right of way and the 2s starting red / amber for the phase gaining right of way. Depending on the geometry of the junction there may also be a period of all red to allow all traffic to clear the junction.  Intergreens must be measured carefully when configuring a junction as one that is too short can mean traffic from the previous stage is still clearing the junction when the next stage starts. Conversely, intergreens which are too long can lead to unnecessary delay within a junction which may have negative implications on the level of queueing and delay experienced at the junction. |
| Interstage Period | The time between the end of one stage and the start of the next stage. |
| Maximum Green | The maximum duration of a green signal after a conflicting demand has been registered in the controller. The value is configurable by time of day to suit different conditions throughout the day. |
| Minimum Green | The minimum duration of a green signal during which time no change of the signal lights can occur. In the UK this is usually 7 seconds for full traffic phases, 4 seconds for an arrow phase and 5 seconds for a pedestrian phase. |
| Invitation Period | The amount of time a steady green signal is shown to pedestrians at traffic signals (either a junction or a standalone pedestrian crossing). |
| Lost Time | The time during a cycle which cannot be used as effective green to one or more phases. |
| Green Split | The division of available green time between stages within a signal cycle. |
| Offset | The time difference between a specific point in the cycle at a junction and a reference point, for example the start / end of a stage at an adjacent junction. |
| Coordination | An arrangement which relates the timings of a signal installation with the timings at a neighbouring installation (junction/crossing etc) |
| Green Wave | A control strategy for a linear system of traffic signals which synchronises the start of green with the arrival of the platoon from the upstream junction. |
| Demand | A request for right of way for traffic. |
| Demand dependent | A stage within a signal cycle that is only called when a demand for it is registered, for example a pedestrian stage or a lightly trafficked side road stage. |
| Call / Cancel | Certain detectors will call a stage or a phase when occupied for a certain period of time. The demand will also be cancelled if the detector then becomes unoccupied for a specified time before the demand has been serviced (for example when a vehicle turns in a gap in opposing traffic). |
| Reserve Capacity | The difference between the capacity of a junction and the current demand. This is usually expressed as a percentage of the current demand. It is used as a measure of how well the junction operates. The higher the reserve capacity (i.e. the more spare capacity) the shorter the queues and delay. |
| Saturation Flow | The maximum flow obtained at a stop line during green from a discharging queue. It is usually expressed in vehicles or passenger car units (pcus) per hour. |
| Oversaturation | Demand exceeding capacity. A measure of how well a junction operates. |
| PCU | Passenger Car Unit. A uniform measurement of vehicles equal to the equivalent of a typical car. A breakdown for all types of vehicle is given in paragraph 3.2 of the TRL publication RR67. |
| Platoon | A group of vehicles moving together where the behaviour of each upstream vehicle is influenced by the one in front. Platoons can often be found progressing around signal controlled roundabouts. |
| Isolated Control | Control of a signal installation where the timings are not related or linked to adjacent installations. |
| Local Control | A method of control for a signal installation which is not influenced by any adjacent junctions or the area control system. |
| MOVA | Microprocessor Optimised Vehicle Actuated method of control for a standalone junction or small group of adjacent junctions which aims to minimise stops and delay and maximise capacity. |
| SCOOT | Split, Cycle, Offset Optimisation Technique which uses real time observed traffic data to minimise stops and delays for UTC controlled areas / networks. |
| UTC | Urban Traffic Control. A method of control for traffic signals where a number of signals are managed by a central computer system. |
| Vehicle Actuation (VA) | A method of control where the green times are determined by the real time flow of traffic on the approaches to the signal installation. Detectors are used to monitor the traffic flows. |
| CLF | Cableless Linking Facility. A method used for coordinating the timings of adjacent signal installations by the use of clocks synchronised to mains electricity supply frequency. |
| Fixed Time | A method of traffic signal control where all the timings are fixed from one cycle to a next i.e. the cycle time, green time and red time all remain constant. |
| Gating | A method of using traffic signals to control the flow of traffic to a downstream junction by restricting it at an upstream junction. |
| SA | Speed Assessment. A control strategy used with VA which takes account of the speed of approaching vehicles when changing the signals. |
| SDE | Speed Discrimination Equipment. A control strategy used with VA which applies specific timings to the junction for vehicles travelling over a given speed. |
| Shuttle Working | A system where traffic signals are used to control traffic over a one way section of road, this can be over a narrow bridge or alongside a section of carriageway that has been narrowed by roadworks. |
| Controller | The apparatus that controls traffic signals. |
| Wig Wag | A signal consisting of two alternately flashing red lights side by side and a single amber light underneath. They are used where vehicles need to stop but where it is difficult for drivers to forecast when this will occur, for example at a swing/lifting bridge, level crossing or by the exit from an emergency vehicle station. Under no circumstances must a vehicle or pedestrian pass through a wig-wag signal that is in operation due to the danger associated with a quickly approaching train, fire engine or opening bridge. |
| Detector | Equipment that initiates, demands or extends on a particular traffic or pedestrian phase. These can either be loops buried in the carriageway or above ground detectors which are often housed on the top of signal poles. |
| Detector Loop | One or more turns of wire installed in the road surface which makes up a detector that relies on the electromagnetic changes of a vehicle passing over it. |
| SVD | Selective Vehicle Detector. A detector which responds to certain vehicles, such as buses or emergency vehicles, either by their characteristics or by an electronic tag. |
| Variable Message Sign | A sign with a message that can be varied to display certain information at certain times for example at different times of day or under different traffic conditions. |

1. Comparative safety of road classes

The following table is taken from RAS10002 from the DfT accident data statistics from 2014 [128]. This compares the reported accident rate for Motorways, Urban Roads and Rural Roads.

Table A3.: Accident Statistics for road types 2014

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Accidents 2014 | Fatal | Fatal and Serious | All Severities | Rate per billion miles |
| Motorways | 85 | 680 | 5630 | 88 |
| All Urban Roads | 591 | 12483 | 95292 | 819 |
| All Rural Roads | 982 | 9171 | 45400 | 340 |
| Totals | 1658 | 22334 | 146322 | 466 |

The information from Table A3.1 can be analysed in many ways. Very simple interpretations are:

* Motorway travel is by far the safest in simple numeric terms. This is attributed to restricted access, restricted vehicle classes, grade segregation of junctions, physical segregation of lanes.
* Accidents are twice as likely to occur on urban roads as they are on rural roads. This is assumed to be an indication of higher density of people, vehicles and street furniture in urban areas.
* Survival of accidents is highest in urban roads (0.62% of accidents included fatalities), followed by motorway (1.51%) and finally rural (2.16%). It is probable that the fatality ratio is inflated on rural roads if the number of minor accidents is significantly under-reported due to inconvenience of reporting and limited effects on other road users. The fatality ratio is lowest on urban routes where impact speeds are likely to be much lower and highest on rural roads where slow moving, unprotected user classes (pedestrians and cyclists) are likely to encounter high speed vehicles.

Other interesting statistics are obtained by calculation. Making the assumption that “All Severities” includes “Fatal and Serious” which includes “Fatal” i.e.

Equ( A3‑‑)

The “All Severities” and “rate per billion miles” can be used to estimate the number of miles driven per road type. This figure is an approximation as the incident rate is rounded to the nearest integer.

Equ( A3‑‑)

The “miles driven” and number of accidents can be used to generate a number to represent the number of miles per incident, again subject to rounding errors.

Equ( A3‑‑)

Table A3.1 and Equ( A3‑0‑1) - Equ( A3‑0‑3) are used to create Table A3.2, which shows the mean distance between incidents of varying severity. The figures presented for Fatal and Fatal and Serious are assumed to be an accurate representation whereas the All Severities figure is very likely to be an overestimate. The reason for this statement is simply that all fatal and serious accidents are reported and attended by emergency services but minor incidents are not. There are therefore likely to be many more unreported incidents.

Table A3.: Incidents per mile 2014

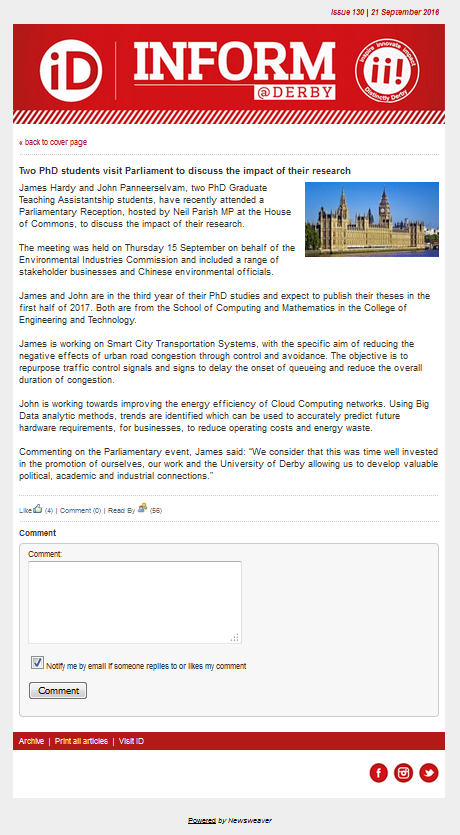
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Accidents 2014 | Miles Driven (billions) | Miles per Fatal incident (millions) | Miles per Fatal and Serious incident (millions) | Miles per All Severities incident(millions) |
| Motorways | 64 | 753 | 94 | 11 |
| All Urban Roads | 116 | 196 | 9 | 1 |
| All Rural Roads | 133 | 135 | 15 | 3 |
| Totals | 365 | 220 | 16 | 2 |

Note: The inconsistent Totals values are mainly attributable to the rounding errors introduced by the rate value in the original figures. The figures are however generally consistent with the values provided in the report “Road Traffic Estimates: Great Britain 2014” [129]. The data shown in Table A3.3 is based on the traffic estimates report. The earlier comment regarding the accuracy of the All Severities data still applies.

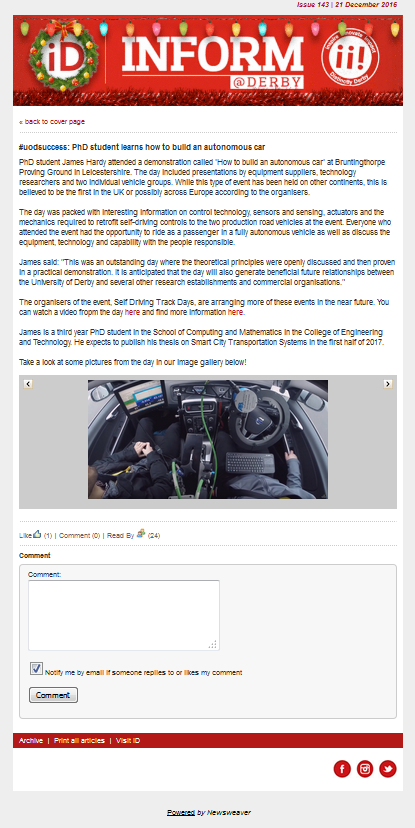
Table A3.: Incidents per mile based on Traffic Estimates report

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Accidents 2014 | Miles Driven (billions) | Miles per Fatal incident (millions) | Miles per Fatal and Serious incident (millions) | Miles per All Severities incident(millions) |
| Motorways | 64.3 | 756.5 | 94.6 | 11.4 |
| All Urban Roads | 114.1 | 193.1 | 9.1 | 1.2 |
| All Rural Roads | 132.5 | 134.9 | 14.4 | 2.9 |
| Totals | 310.9 | 187.5 | 13.9 | 2.1 |

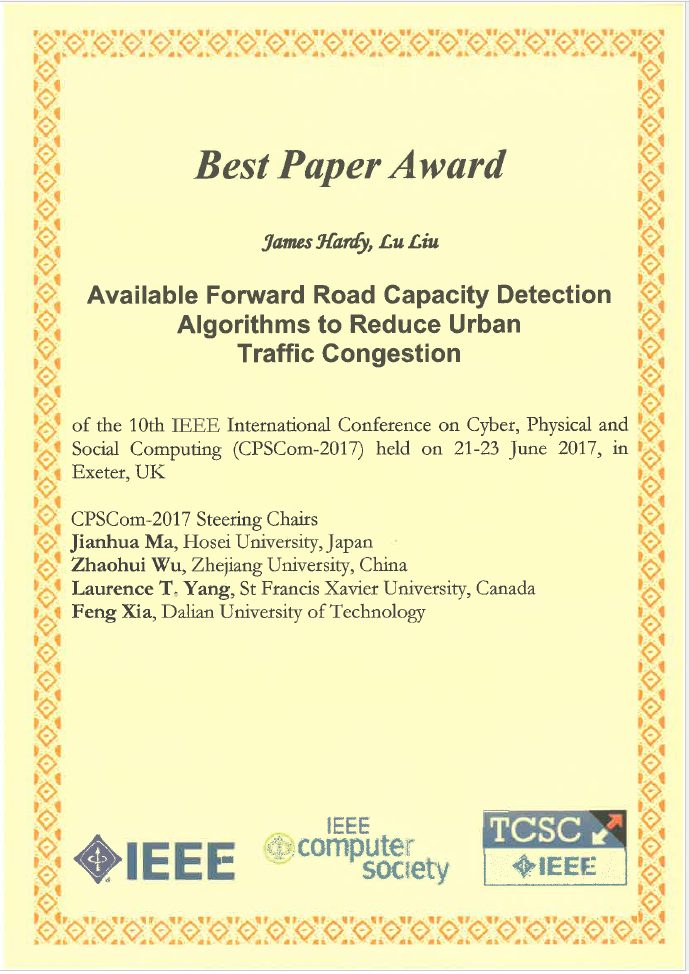
1. Visit to Parliament



1. Autonomous vehicle testing visit



1. Best Paper Certificate



1. Sample database outputs



